Molecular characterisation of the UgpB substrate binding protein from *Mycobacterium tuberculosis*

Jonathan Fenn

Supervised by

Dr Elizabeth Fullam and Professor Alex Cameron

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Interdisciplinary Biomedical Research

University of Warwick

School of Life Sciences and Warwick Medical School

MRC Doctoral Training Programme

September 2019

Contents

Table of figures

Table of tables

Acknowledgements

First and foremost I would like to thank my supervisors, Dr Elizabeth Fullam and Professor Alex Cameron for supporting me over the past 3 years during my PhD. I would like to express my gratitude to Liz for her continual support, guidance and constructive criticisms throughout the PhD. I have always felt encouraged and I believe I have developed many skills from being a student under her supervision that I value and will continue to draw upon in the future. I express my thanks to Alex especially for helping me to solve the X-ray structures and for experimental advice.

I express my gratitude to the MRC Doctoral Training Programme Directors, Professor Andrew McAinsh, Professor Jonathan Millar and all members of the leadership team for their support and guidance and for accepting me onto the PhD programme. I especially thank Sally Blakeman for organising all of the enjoyable student conferences and always being there for any queries. I thank the Medical Research Council for funding my studentship. I also want to thank Professor Mohan Balasubramanian and Dr Saravanan Palani for inspiring me to do a PhD during my undergraduate project.

I also thank everybody in the lab, past and present members. Especially Dr Collette Guy for helping me with the chemistry experiments and characterisations, Dr Chris Furze for guidance in biochemistry and molecular biology. Dr Mohd Syed Ahanger for helping me with crystallography and at the Diamond Light Source, Dr James Harrison for helping me with MST, Magda, Marie, Hadyn and all members of the group for their helpful discussions and for making my PhD such an enjoyable experience. I will especially remember all of the World Tuberculosis Day events we held together, the Astbury Conversation and the Acid Fast Club Meetings we attended.

I also thank our collaborators Dr Ridvan Nepravishta and Dr Jesus Angulo from the University of East Anglia for carrying out the STD-NMR experiments and for helping to progress my project.

I also thank Professor Matthew Gibson for allowing me to work in the chemistry lab and for guidance with the different synthesis strategies. Dr Sarah-Jane Richards for helping me use the Octet instrument and experimental assistance. The mass spectrometry and NMR service for processing my samples. Thanks also to the Warwick Integrative Synthetic Biology facility for providing access to the MST instrument and the proteomics facility, Dr Juan Hernandez-Fernaud and Dr Cleidiane Zampronio for experimental assistance.

Lastly, I really appreciate my girlfriend Becky, my friends and parents for being there and for all of their support, encouragement and advice during my PhD.

Declaration

The work presented in this thesis is all the original work of the author except where it contains work based on collaborative research, in which case the nature and extent of the author's individual contribution is clearly indicated. I confirm that the work presented in this thesis has not been submitted for a degree at another university.

Publications

Work from Chapters 2, 3, 4 and 5 have been published as:

Fenn, J.S., Nepravishta, R., Guy, C.S., Harrison, J., Angulo, J., Cameron, A.D. and Fullam, E., (2019). Structural basis of glycerophosphodiester recognition by the *Mycobacterium tuberculosis* substrate-binding protein UgpB. *ACS Chem. Biol.* 2019 14(9),1879-1887, DOI: 10.1021/acschembio.9b00204

Conference proceedings:

Fenn, J.S., (2018). Structural and functional analysis of the solute-binding protein UgpB involved in glycolipid recognition from *Mycobacterium tuberculosis.* Awarded poster prize: Astbury Conversation, University of Leeds.

Fenn, J.S., (2018). Structural and functional analysis of the UgpABCE transporter of *Mycobacterium tuberculosis*. Awarded prize for best Year 2 PhD presentation: MRC DTP in Interdisciplinary Biomedical Research Student Conference, Wilderhope Manor, Shropshire.

MRC DTP in Interdisciplinary Biomedical Research Student Conference, Nightingale Centre, Derbyshire (2019).

Postgraduate Research Degree Student Symposium, Warwick Medical School (2019).

Joint King's College London and Warwick MRC DTP Conference (2018).

MRC DTP in Interdisciplinary Biomedical Research Student Conference, Hartington Hall, Derbyshire (2017).

Abstract

Mycobacterium tuberculosis (*Mtb*) is the causative agent of tuberculosis (TB) and remains as one of the leading causes of death worldwide. *Mtb* encodes a limited number of ATP-binding cassette (ABC) importers for the acquisition of carbohydrates, potentially reflecting the nutrient scarce environment encountered within the host. The focus of this PhD therefore, was to carry out a structural and functional analysis of *Mtb* UgpB (Rv2833c), the substrate-binding protein of the *Mtb* UgpABCE transporter.

The *Mtb* UgpB substrate-binding protein was successfully expressed and purified to homogeneity. Following purification, the function of *Mtb* UgpB was investigated. A panel of potential substrates were investigated using a thermal shift assay, which found that glycerol-3-phosphocholine (GPC) produced the highest thermal shift. Microscale thermophoresis revealed a clear binding preference for GPC with a K_d of 3.6 μM for apo-*Mtb* UgpB. Further to these studies, alternative glycerophosphodiesters were also found to bind to *Mtb* UgpB, demonstrating that *Mtb* UgpB is promiscuous for a wide range of glycerophosphodiester substrates. Structural studies were undertaken in both solid and solution state and revealed, for the first time, the molecular basis of GPC and glycerophosphodiester recognition. Genetically engineered variants of *Mtb* UgpB were created to investigate the direct role of residues that interact with GPC and further define the molecular basis of substrate recognition and the specificity for glycerophosphodiesters. In addition, further studies were undertaken to produce additional components of the *Mtb* UgpABCE transporter and the nucleotide binding protein, *Mtb* UgpC (Rv2832c) was successfully expressed and its ATPase activity investigated.

Combined, these results have provided critical insights into the structural and functional role of *Mtb* UgpB, revealing the specificity is not limited to GPC, and that *Mtb* UgpB may have adapted to scavenge alternative glycerophosphodiester metabolites *via* a single transporter during intracellular infection.

Abbreviations

3D	Three dimensional
ABC	ATP-binding cassette
ADP	Adenosine diphosphate
AG	Arabinogalactan
AIDS	Acquired immune deficiency syndrome
ATP	Adenosine triphosphate
BCG	Bacillus Calmette-Guerin
BLI	Biolayer interferometry
BSA	Bovine serum albumin
DAT	Diacyl trehalose
DEEP	Differential epitope mapping
DEFRA	Department for Environment, Food and Rural Affairs
DMAB	Borane dimethylamine complex
DNA	Deoxyribonucleic acid
DprE1	Decaprenylphosphoryl-D-ribose-2-epimerase
ECL	Enhanced chemiluminescence
EM	Electron microscopy
FDA	Food and Drug Administration
FRET	Forster resonance energy transfer
G2P	Glycerol-2-phosphate
G3P	Glycerol-3-phosphate
GFP	Green fluorescent protein
GPC	Glycerol-3-phosphocholine
GPE	Glycerophosphoethanolamine
GPG	Glycerophosphoglycerol
GPI	Glycerophosphoinositol
GPI(4)P	Glycerophosphoinositol-4-phosphate
GPS	Glycerophosphoserine
GST	Glutathione S-transferase
HIV	Human immunodeficiency virus
ID	Identity
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl-b-D-thiogalactoside
ITC	Isothermal calorimetry
K_d	Binding affinity

LAM	Lipoarabinomannan
LED	Light emitting diode
LM	Lipomannan
mAGP	Mycolyl-arabinogalactan-peptidoglycan
MBP	Maltose binding protein
MDR	Multi drug resistant
MFS	Major facilitator superfamily
MRC	Medical research council
MSD	Membrane spanning domain
MST	Microscale thermophoresis
Mtb	*Mycobacterium tuberculosis*
MTBC	*Mycobacterium tuberculosis* complex
NBP	Nucleotide binding protein
NMR	Nuclear magnetic resonance
ORF	Open reading frame
PBS	Phosphate buffered saline
PC	Phosphatidylcholine
PCR	Polymerase chain reaction
PDB	Protein data bank
PDIM	Phthiocerol dimycocerosates
PEG	Polyethylene glycol
PG	Peptidoglycan
P_i	Inorganic phosphate
PI	Phosphatidylinositol
PIM	Phosphatidylinositol mannoside
PS	Phosphatidylserine
RD	Region of difference
RNA	Ribonucleic acid
RR	Rifampicin resistant
SBP	Substrate binding protein
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate Polyacrylamide gel electrophoresis
SIV	Simian immunodeficiency virus
SL	Sulfolipid
SNP	Single nucleotide polymorphism
SPR	Surface plasmon resonance
STD	Saturation transfer difference

SUMO	Small ubiquitin-like modifier
TAG	Triacylglycerol
TB	Tuberculosis
TBS	Tris buffered saline
TDM	Trehalose dimycolate
TDR	Totally drug resistant
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
T_m	Thermal shift
TMM	Trehalose monomycolate
TNF	Tumour necrosis factor
UN	United Nations
WHO	World Health Organisation
XDR	Extensively drug resistant

Chapter 1: Introduction

1.1 Introduction

1.1.1 Tuberculosis history

Tuberculosis (TB) has a vast history of infecting humans spanning over thousands of years (Morse et al., 1964). Paleopathology and ancient DNA analyses have revealed that TB was widespread in ancient Egypt (Crubezy et al., 1998) as Pott's lesions that are characteristic of spinal TB have been identified in Egyptian mummies dating back to 2400 BC (Zimmerman, 1979). There is also evidence of *Mycobacterium tuberculosis* (*Mtb*) complex DNA and cell-wall mycolic acids identified in the first Egyptian mummy to be scientifically studied, "Dr Granville's" mummy, confirming the likely cause of death to be TB (Donoghue et al., 2010). Amazingly TB was also present in South America long before the first European's had colonised as archaeological and DNA evidence have been identified in the remains of ancient Peruvian mummies (Salo et al., 1994), indicating that TB was globally spread in ancient times just as it remains today.

At around the start of the industrial revolution in Europe by the late 18th and early 19th Centuries, TB was an epidemic and staggeringly the leading cause of death in many counties (Barberis et al., 2017). It was reported there were 800-1,000 deaths per 100,000 of population per year (Krause, 1928). During this period of history the basis of infectious diseases were being studied extensively and the causative agent of tuberculosis, the bacillus *Mtb* was first isolated by Dr Robert Koch in 1882 (R. Koch, 1882). However TB couldn't be effectively treated and it wasn't until the first half of the 20th Century when the first antibiotics were discovered and developed to successfully treat tuberculosis (Schatz et al., 1944) (Lehmann, 1946) (Robitzek et al., 1952) . Streptomycin was the first anti-TB antibiotic to be developed. Initially it appeared to reduce TB mortality (Long et al., 1950) however drug resistance developed rapidly and it was later found to be ineffective if used in isolation (Fox et al., 1954). As new drugs were discovered and developed for TB, combination therapies were investigated due to the possibility of reducing relapse/drug resistance (BMJ, 1955). It was later found that combining streptomycin and the first synthetic anti-TB drug, *para*-aminosalicylic acid reduced the incidence of streptomycin resistance however it didn't completely resolve the resistance problem (Fox et al., 1956). A new drug: isoniazid, was discovered in 1952 and found to be extremely effective; however, when used alone, resistance also occurred in 70 % of cases after 3 months of treatment (Sita Lumsden et al., 1952). As almost all of the identified resistant strains were only resistant to one of the three available drugs this led to increased investigations into three drug regimens (Fox et al.,

1955). The first curative regime is considered to be isoniazid combined with streptomycin and *para*-aminosalicylic acid and it had very low relapse rates of only 4 % for people treated for 1-2 years (MRC, 1962) (Fox et al., 1999). However more effective antibiotics to treat TB were later developed. Pyrazinamide was discovered and found to kill the *Mtb* bacilli that persisted after treatment with isoniazid and streptomycin (R. M. McCune, Jr. et al., 1956) and rifampicin was found to increase killing of the *Mtb* bacilli located in mouse organs (Grumbach, 1970). A compound far more potent than streptomycin was also discovered called ethambutol (Thomas et al., 1961). The discovery of these more effective drugs led to a continued focus on the ease of administration and reducing the treatment duration and adverse side effects. These developments eventually led to the current recommended 6 month regimen (WHO, 2010) for treating drug susceptible TB: A 2-month intensive phase of isoniazid, rifampicin, pyrazinamide and ethambutol combined, followed by a continuation phase of isoniazid and rifampicin for 4 months.

Since the implementation of drug treatments developed from 1940-1960 there had been a gradual decline in the number of new cases of TB and mortality rates in the developed world (Griffith et al., 1996). TB was beginning to be controlled and considered a disease of the past (Griffith et al., 1996). However the incidence of TB increased dramatically during the 1980's to 1990's, mainly fuelled by the HIV/AIDS epidemic (Havlir et al., 1999) and inadequate treatment practices leading to the development of multi-drug resistant TB (Frieden et al., 1993) (Seung et al., 2004). Now, in 2019 according to the latest report from the World Health Organisation (WHO, 2018), TB is one of the top 10 leading causes of death worldwide and in 2017 an estimated 1.7 million people died and 10.0 million new people acquired the disease mostly in less developed countries (WHO, 2018). Disease incidence varies hugely between different counties (Figure 1.1) and TB is generally associated with poverty and overcrowding (Hill et al., 2006). Developed countries have <10 new cases per 100,000 people (WHO, 2018). In contrast, the 30 highest TB burden countries have 150-400 new cases per 100,000 and the 3 countries with the highest burden of TB: Mozambique, Philippines and South Africa have >500 new cases per 100,000 (WHO, 2018). Shockingly, over the past 200 years it's estimated that globally 1 billion people have died as a result of TB, more than from malaria, influenza, smallpox, HIV/AIDS, cholera and the plague combined (Paulson, 2013).

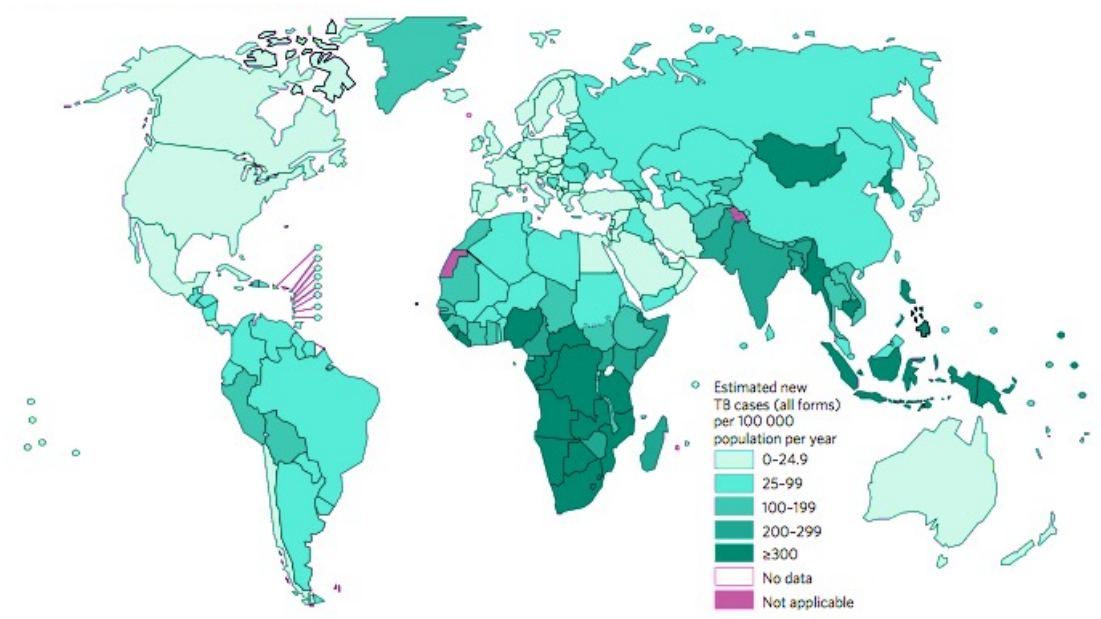


Figure 1.1: Global incidence map of tuberculosis in 2017 (WHO, 2018)**.** The highest incidences of TB appear to be in Sub-Saharan Africa and South East Asia. Shaded colours represent the estimated number of new TB cases per 100,000 population per year.

1.1.2 Drug resistant TB

Currently drug resistant TB is a looming public health crisis and it's estimated that in 2017, 558,000 people developed rifampicin resistant TB (RR-TB) (WHO, 2018). Rifampicin is one of the most effective frontline anti-TB drugs (Goldstein, 2014). Of the 558,000 RR-TB individuals, 82 % had developed multidrug resistant TB (MDR-TB), defined as resistance to both rifampicin and isoniazid. This trend appears to be being driven by poor treatment practices and lack of drug availability (WHO, 2018). Astonishingly almost half of the cases of MDR-TB were attributed to only 3 countries combined: India, China and Russia. MDR-TB is much more difficult to treat and as a consequence resistance appears to be getting worse as 8.5 % of drug resistant TB is now extensively drug resistant (XDR-TB) - defined as being resistant to rifampicin, isoniazid and the second-line anti-TB drugs including fluoroquinolones and one of the injectable drugs (Table 1.1). Furthermore there are reports of a small number of cases of totally drug resistant (TDR-TB) for which there is no cure (Udwadia et al., 2012). One of the main problems is that there are not enough new drugs being developed to meet the drug resistance crisis (Laughon et al., 2017) as pharmaceutical companies do not have profit incentives to invest. There is however some positive news as recently the FDA approved a new more effective and shorter, 6-month treatment for MDR-TB and XDR-TB using pretomanid combined with bedaquiline and linezolid (FDA, 2019).

Table 1.1: Current first line anti-TB drugs and second line anti-TB drugs used to treat rifampicin resistant (RR-TB) and multi drug resistant tuberculosis (MDR-TB). Table adapted from (WHO treatment guidelines for drug-resistant tuberculosis, 2016).

Grouping of drugs	Drug
First line anti-TB drugs	Isoniazid, rifampicin, ethambutol, pyrazinamide
Second line anti-TB drugs Recommended for the treatment of rifampicin-resistant and multidrug-resistant tuberculosis.	During the 8 month intensive phase: Pyrazinamide + 4 second line drugs: 1 (group A), 1 (group B), at least 2 (group C), a drug from group D2 and D3 if the combination cannot be composed from the above.
Group A	Fluoroquinolones: (levofloxacin, moxifloxacin, gatifloxacin)
Group B	Injectables: (amikacin, capreomycin, kanamycin, streptomycin)
Group C	Other agents: (ethionamide, cycloserine, linezolid, clofazimine
Group D	Add-on agents: (D1: pyrazinamide, ethambutol, high-dose isoniazid, D2: bedaquiline, delamanid, D3: *p*-aminosalicylic acid, imipenem, meropenem, amoxicillin, thioacetazone)

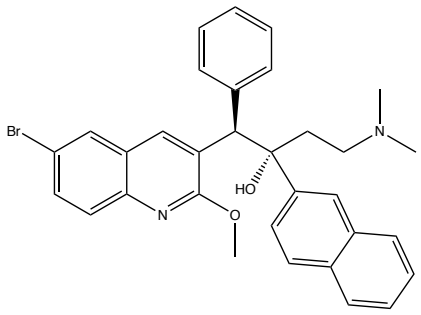
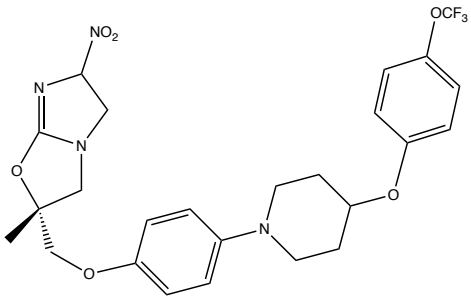
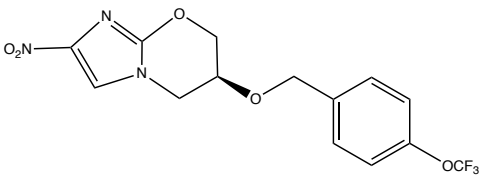
1.1.3 WHO End TB Strategy

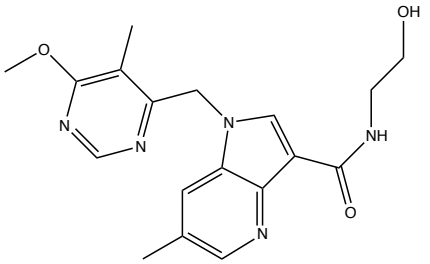
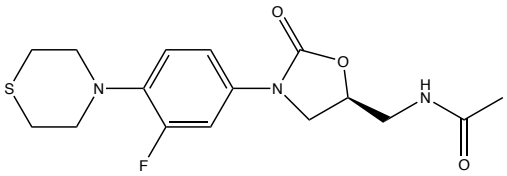
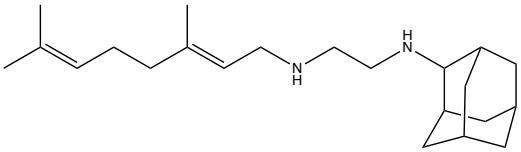
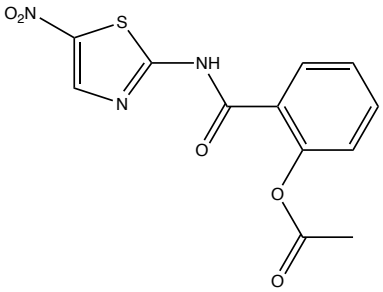
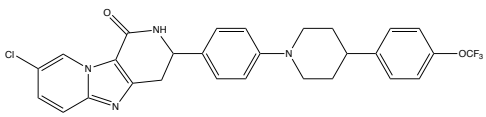
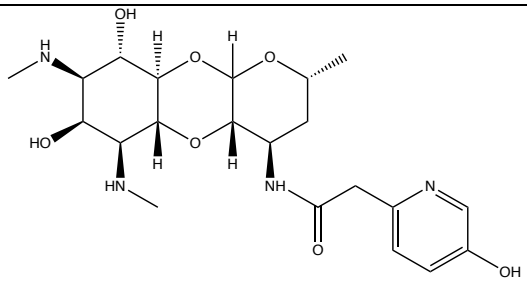
The WHO has bold aims to tackle TB and in 2014 and 2015 all member states of the WHO and the United Nations (UN) made a commitment to end the TB epidemic by endorsing the End TB strategy (WHO, 2014). The End TB Strategy has set out key milestones in 2020 and 2025 to reduce the number of TB cases and deaths in order to end the TB epidemic by 2035 (WHO, 2014). The specific target for the End TB Strategy is for a 90 % reduction in the number of TB deaths and 80 % reduction in TB incidence compared with levels in 2015. However, currently the worldwide TB disease burden is not falling fast enough to meet the first 2020 milestone of a 35 % reduction in the number of TB deaths and 20 % reduction in incidence (WHO, 2018). Better treatments and diagnostics are needed however this requires a more extensive fundamental understanding of the pathophysiology of TB.

1.1.4 Anti-TB drugs in the clinical pipeline

Due to the emergence of drug resistant TB, treatment regimens have had to be altered to accommodate different drugs and treatment durations have had to be increased. Many of the current anti-TB drugs have adverse side effects therefore there is an urgent need to develop new drugs that are safer and more effective and require shorter treatment durations. There are several new drugs currently in clinical development (Table 1.2).

Table 1.2: Anti-TB drugs currently in clinical development. Adapted from (Tornheim et al., 2019) (Libardo et al., 2018) and TB Alliance.

Anti-TB drug	Structure	Class	Target
Bedaquiline FDA approved MDR-TB Phase III		Diarylquinoline	Binds to *Mtb* ATP synthase and inhibits respiration
Delamanid FDA approved MDR-TB Phase III		Nitroimidazole	Blocks synthesis of mycolic acids and forms nitrogen reactive species
Pretomanid FDA approved MDR-TB and XDR-TB Phase III		Nitroimidazole	Blocks synthesis of mycolic acids and forms nitrogen reactive species

TBA-7371 Phase II		DprE1 inhibitor	Decaprenylphosphoryl-β-D-ribose 2'epimerase
Sutezolid Phase II		Oxaziolidinone	N-formylmethionyl-tRNA binding to the ribosome
SQ109 Phase II		Ethylenediamine	Binds to MmpL3 and inhibits cell wall synthesis
Nitazoxanide Phase II		Nitrothiazole	Disrupts membrane potential and pH homeostasis
Q203 Phase I/II		Imidazopyridine	Binds to the QcrB subunit of cytochrome c1 and inhibits respiration
Spectinamide Pre-phase I		Spectinamide	Binds to the ribosome and inhibits protein synthesis

Bedaquiline is a new diarylquinoline anti-TB drug and is the first in its class, currently FDA approved for the treatment of MDR-TB (Andries et al., 2005). Bedaquiline is the first anti-TB drug that binds to *Mtb* ATP synthase and inhibits cellular respiration (Preiss et al., 2015). Delamanid is a bicyclic nitroimidazole and is currently FDA approved for the treatment of MDR-TB. Delamanid inhibits synthesis of mycolic acids, which is expected to enhance *Mtb* drug susceptibility although the exact target is not currently known (Manjunatha et al., 2009). Pretomanid is also a nitroimidazole and was recently approved for the treatment of MDR-TB and XDR-TB in combination with bedaquiline and linezolid (FDA, 2019). TBA-7371 is a potent anti-TB compound that targets *Mycobacteria* decaprenylphosphoryl-D-ribose-2-epimerase (DprE1) by non-covalent inhibition (Gawad et al., 2018). DprE1 is an epimerase involved in a crucial step of *Mycobacterial* cell wall lipoarabinomannan and arabinogalactan biosynthesis where decaprenylphosphoryl-D-ribose is converted to decaprenylphosphoryl-D-arabinose (Gawad et al., 2018). TBA-7371 is currently in phase II clinical trials and has demonstrated *in vivo* efficacy and safety (TB Alliance). Sutezolid is a oxaziolidinone analogue of the currently approved anti-TB drug, linezolid (Maartens et al., 2015). Sutezolid targets binding of N-formylmethionyl tRNA to the ribosome and inhibits bacterial protein synthesis (Wallis et al., 2014). Sutezolid is currently in phase I clinical trials and research has found that it may be clinically more effective compared to linezolid (TB Alliance). SQ109 is an ethylenediamine, related to ethambutol and is currently in phase II clinical trials. SQ109 was found to disrupt mycolic acid cell wall assembly by potentially targeting MmpL3, the transporter of trehalose monomycolate (TMM) (Tahlan et al., 2012). Nitazoxanide is an approved anti parasitic drug however it was found to have bactericidal activity against *Mtb* and is currently being tested in phase II clinical trials (Shigyo et al., 2013). Q203 is an imidazopyridine that targets respiration by inhibiting the QcrB subunit of the cytochrome. Q203 has been shown to inhibit MDR and XDR-TB growth (Pethe et al., 2013). Spectinamide is in phase I clinical trials and is being developed to treat drug resistant TB. Spectinamide inhibits the ribosome (Liu et al., 2017) and can overcome efflux by the *Mtb* Rv1258c efflux pump. Many of these compounds have novel activities; however considering the possibility of failure there is still an urgent need to develop new drugs.

1.1.5 Causative agent of tuberculosis

Dr Robert Koch first isolated *Mtb* the causative agent of TB in 1882. *Mtb* is part of the *Mycobacteria* genus made up of approximately 170 species many of which are non-pathogenic, environmental saprophytes commonly found in the soil and water (Fedrizzi et al., 2017). *Mycobacteria* species have a characteristic morphology being rod-shaped, acid-fast bacilli and have a unique thick and waxy cell wall (Petit et al., 1969)

(Lederer et al., 1975) (Minnikin et al., 1984) (Besra & Brennan, 1997) (Alderwick et al., 2015). Historically the different *Mycobacteria* species have been classified as being fast growers or slow growers however they can be classified based on differences in 16S rRNA sequence (Figure 1.2). Interestingly all of the human pathogens are found in the slow-growing group (Rogall et al., 1990).

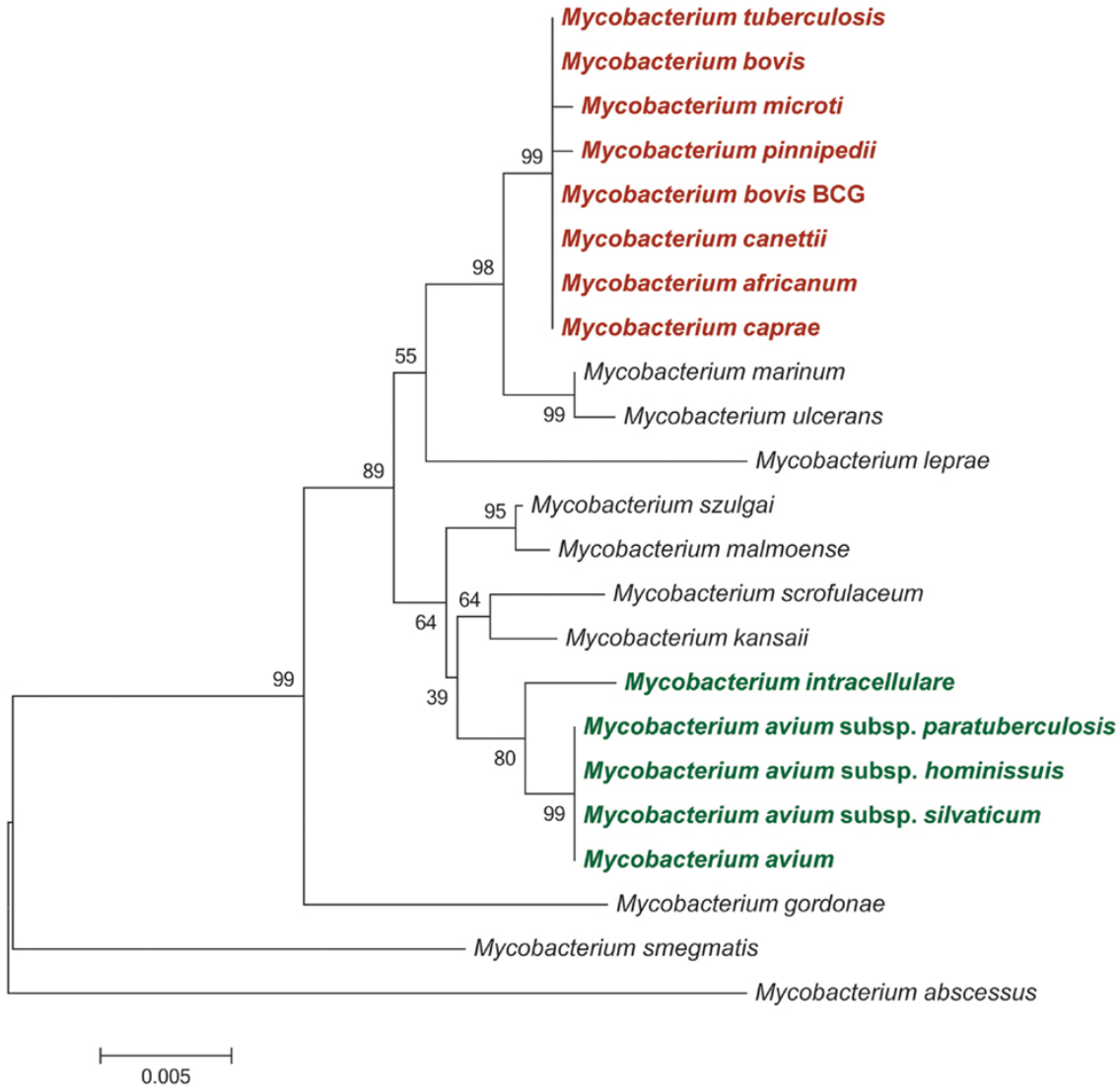


Figure 1.2: Phylogenetic tree of selected *Mycobacteria* species based on partial 16S rRNA sequence (Rue-Albrecht et al., 2014). Red refers to species of the *Mycobacterium tuberculosis* complex MTBC and green refers to species of the *Mycobacterium avium* complex.

1.1.6 The *Mycobacterium tuberculosis* complex (MTBC)

The *Mycobacterium tuberculosis* complex (MTBC) consists of species that can cause tuberculosis. Members of the MTBC include: *Mycobacterium tuberculosis, Mycobacterium africanum, Mycobacterium bovis, Mycobacterium microti, Mycobacterium canettii, Mycobacterium pinnipedii, Mycobacterium caprae, Mycobacterium orygis, Mycobacterium suricattae* and *Mycobacterium mungi*

(Gagneux, 2018). Species of the MTBC are very closely related at the nucleotide level sharing >99 % sequence similarity (Brosch et al., 2002) apart from *Mycobacterium canettii* that is only 97.3 % similar (Supply et al., 2013). They are most effectively distinguished by genetic markers including single nucleotide polymorphisms (SNP's) in the 16S rRNA gene, DNA gyrase subunit B *gyrB* gene and large sequence polymorphisms called regions of difference (RD) (Huard et al., 2006). Other ways of distinguishing the different species are by their characteristic phenotypes, different host niches and pathogenicity.

It was first believed that transmission of *Mycobacterium bovis* from domesticated animals during the Neolithic age led to the evolution of *Mtb* and human's acquiring tuberculosis however more recent evidence suggests this is not the case (Manchester, 1984). For example the order of gene loss identified from the *Mycobacterium bovis* genome (Garnier et al., 2003) suggested that humans may have actually transmitted tuberculosis to animals (Brosch et al., 2002) (Mostowy et al., 2002). It is currently believed that the MTBC complex evolved from environmental *Mycobacteria* that adapted and gained the ability to survive intracellularly. One theory is that amoeba, a type of free-living animal more accurately a protozoon, that normally consumes bacteria could have provided the intracellular niche for the MTBC progenitor to first evolve (Jang et al., 2008). The first MTBC progenitor is likely to have acquired genes potentially by horizontal gene transfer and also gene loss that enabled it to survive inside anaerobic conditions however no single genetic feature appears to be responsible for MTBC virulence (Veyrier et al., 2011).

More recently evidence seems to point to the origin of the human adapted MTBC being from Africa, and Africa incidentally also has all 7 human adapted MTBC lineages (Figure 1.3) and the largest lineage diversity (Gutierrez et al., 2005). It is thought that the first MTBC progenitor evolved from an organism most similar to *Mycobacterium canetti* (Supply et al., 2013), a smooth tubercle bacilli that are the closest relatives of the MTBC only found in the Horn of Africa and sharing 97.3 % nucleotide similarity to *Mtb*. However the exact age of the MTBC progenitor is not known, one study estimated it to have existed 70,000 years ago based on the co-divergence of humans and MTBC and the first migrations out of Africa (Comas et al., 2013).

Of the MTBC species, only two are primarily associated with causing tuberculosis in humans and include *Mtb* and *Mycobacterium africanum* (Brites et al., 2015). *Mtb* and *Mycobacterium africanum* are further divided into 7 phylogenetic lineages and have enabled ecological mapping: L1, L2, L3, L4 and L7 are the *Mtb* lineages and L5 and L6

are the *Mycobacterium africanum* West Africa 1 and 2 lineages. *Mtb* L2 and L4 lineages are the most wide spread lineages throughout the world (Figure 1.3) and cases of TB have been reported on every continent.

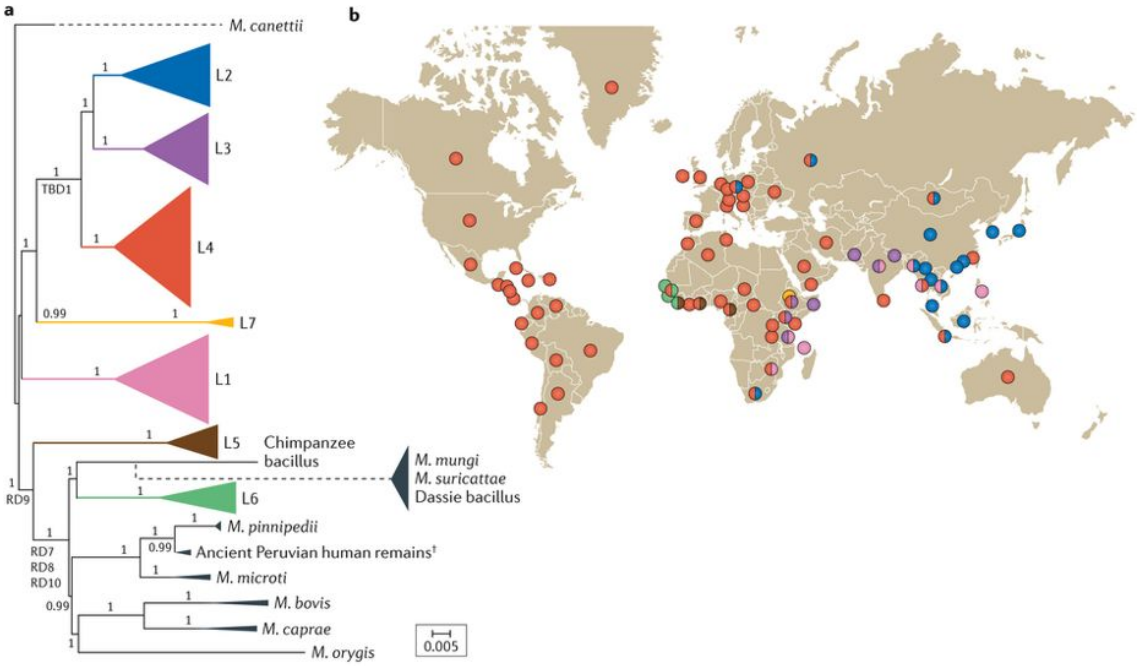


Figure 1.3: Genetic phylogeny and global distribution of the MTBC lineages (Gagneux, 2018). Coloured symbols refer to the seven human-adapted MTBC lineages. Grey symbols refer to the primarily animal adapted MTBC lineages. RD refers to genome region of deletion found in the lineages. *Mycobacterium tuberculosis* lineages L2 and L4 are the most wide spread globally.

The other MTBC species are mainly adapted to cause tuberculosis in animals; more specifically mammals however there have also been reports of transmission events to humans. For example *Mycobacterium bovis* is the primary cause of tuberculosis in cattle and believed to be transmitted from wildlife animals where it causes significant economic damage to the agriculture industry (Pollock et al., 2002). In the United Kingdom, the role that badgers have in maintaining bovine tuberculosis in cattle is very controversial. The results of the randomised badger culling trial found that badger culling reduced the incidence of TB in cattle herds (Donnelly et al., 2007). However if less than 70 % of badgers are culled in the area defined by the study there is evidence that bovine TB could increase due to the perturbation effect (Donnelly et al., 2006). In the long-term badger culling is not considered to be sustainable and it is also strongly opposed by the public for being inhumane. Furthermore the validity of DEFRA published data (DEFRA, 2014) has been strongly criticised for being misleading and puts into question the efficacy of these culls. Interestingly there have been cases of humans acquiring *Mycobacterium bovis* and developing tuberculosis (Grange, 2001) and the consumption of unpasteurised milk or products made from infected animals is

thought to be how it is transmitted (Hancox, 2002). More recently transmission of *Mycobacterium pinnipedii* from sea lions to humans in a zoo was reported to have caused tuberculosis in humans therefore there appears to be the propensity for other members of the MTBC to cross the species barrier and infect humans (Kiers et al., 2008). In contrast other non-tuberculosis *Mycobacteria* species exist that infect humans however do not cause tuberculosis including *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium marinum Mycobacterium xenopi*, *Mycobacterium gordonae* and *Mycobacterium kanasii* (Gagneux, 2018). Genetically these species have diverged from the MTBC (Porvaznik et al., 2017) for example *Mycobacterium avium* is associated with causing disease mostly in immunocompromised individuals (Kiehn et al., 1985) and *Mycobacterium marinum* is able to cause skin infections and tuberculosis-like disease in immunocompromised individuals (Aubry et al., 2002) (Parent et al., 1995).

1.1.7 Comparative genomics of the *M. tuberculosis* complex (MTBC)

The complete genome sequence of the first characterised strain of *Mtb*, H37Rv was deciphered in 1998 (Cole et al., 1998) however many other strains of *Mtb* have since been fully sequenced (Zhang et al., 2011) (Manson et al., 2017). The complete genome sequence revealed that the *Mtb* genome is approximately 4.4 Mbp and has a very high guanine + cytosine content. Currently the *Mtb* genome is annotated to encode 4173 genes (Mycobrowser, 2019). Several other species of *Mycobacteria* complete genomes have since been sequenced and are available (Sanger, 2019) These datasets are invaluable tools to investigating the fundamental biology of *Mycobacteria* because comparative genomics can be used to explore differences at the genetic and biochemical level that may influence pathogenicity and survival. One of the key insights came from comparing the genomes of *Mycobacterium tuberculosis*, to *Mycobacterium bovis* and *Mycobacterium bovis* BCG (Mahairas et al., 1996), the live-attenuated vaccine strain that is unable to infect humans. Comparative genomics revealed 16 distinct genomic regions of deletion that were deleted in *Mycobacterium bovis* and *Mycobacterium bovis* BCG (Behr et al., 1999) that may have a role in its loss of pathogenicity. Of particular note was a spontaneous deletion in the type VII secretion system ESX-1 a key virulence factor involved in host cell lysis that led toward the attenuation of *Mycobacterium bovis* BCG (Groschel et al., 2016).

Another important genomic insight came from sequencing the complete genome of *Mycobacterium leprae* (Vissa et al., 2001). *Mycobacterium leprae* is the causative agent of leprosy (R. Koch, 1882). This study revealed that *Mycobacterium leprae* had undergone massive genome decay and pseudogenes now occupy half of its genome

(Vissa et al., 2001). The *Mycobacterium leprae* genome is only 3.2 Mbp and approximately one third of its genome had been lost. Because *Mycobacterium leprae* is pathogenic, its 2963 genes (Mycobrowser, 2019) are considered to be the minimal set of genes required for the survival and pathogenesis of *Mycobacteria* (Monot et al., 2009). Therefore any genes conserved between *Mycobacterium tuberculosis* and *Mycobacterium leprae* are likely to be essential for its survival (Vissa et al., 2001).

1.1.8 Finding essential genes

Complete genome sequencing of *Mycobacteria* species has led to advancements in identifying the genes that are essential for *Mtb* growth, that are therefore most likely to be suitable drug targets. Early technologies used transposon site hybridisation (TraSH) knockout libraries and microarrays (Sassetti et al., 2001) to identify essential genes however deep sequencing of transposon (Tn) insertion (TnSeq) libraries has been a more powerful tool to identify essential genes required for *Mtb* growth *in vitro* and in host macrophages *in vivo.* The most recent study identified 625 open reading frames (ORF's) that are essential for optimal *Mtb* growth *in vitro* (DeJesus et al., 2017). A drawback of these studies is that they do not fully reflect the nutrient scarce environment encountered inside host macrophages.

1.1.9 Lifecycle and pathogenesis of *M. tuberculosis* infection

Mtb is well adapted to infect humans. However, unlike many other pathogens, due to its infection life cycle it has to cause disease in order to transmit between individuals (Brites et al., 2012). TB primarily affects the lungs where it is called pulmonary TB however the bacteria have the potential to enter the lymphatic system and bloodstream, disseminate to distant tissue and cause disease in TB susceptible regions including the lymph nodes, kidneys, brain, bones and spine (Yates et al., 1993) (Garg et al., 2011) (Eastwood et al., 2001) (Garg, 1999). Once disseminated the disease is now classified as extra-pulmonary TB, and is more difficult to treat (J. Lee, 2015). For the majority of individuals that are initially infected with *Mtb*, the immune response is sufficient to effectively control the infection however individuals with weakened immune systems are inherently more susceptible to developing active TB (Cadena et al., 2017). Individuals with active TB are clinically symptomatic and able to readily transmit the disease. However for the other infected individuals classified as latent TB infected, containment of the disease is a lifelong standoff where bacterial replication is controlled by the immune system yet there is the potential for it to reactivate and become active TB (Getahun et al., 2015). Approximately one-quarter of the global population are latent infected (WHO, 2018). This represents a huge potential infection reservoir of clinically asymptomatic, non-infectious individuals and

approximately 5-10 % of latent infected TB individuals will develop active TB at some point during their lifetime (Figure 1.4) (Koul et al., 2011) (WHO, 2018).

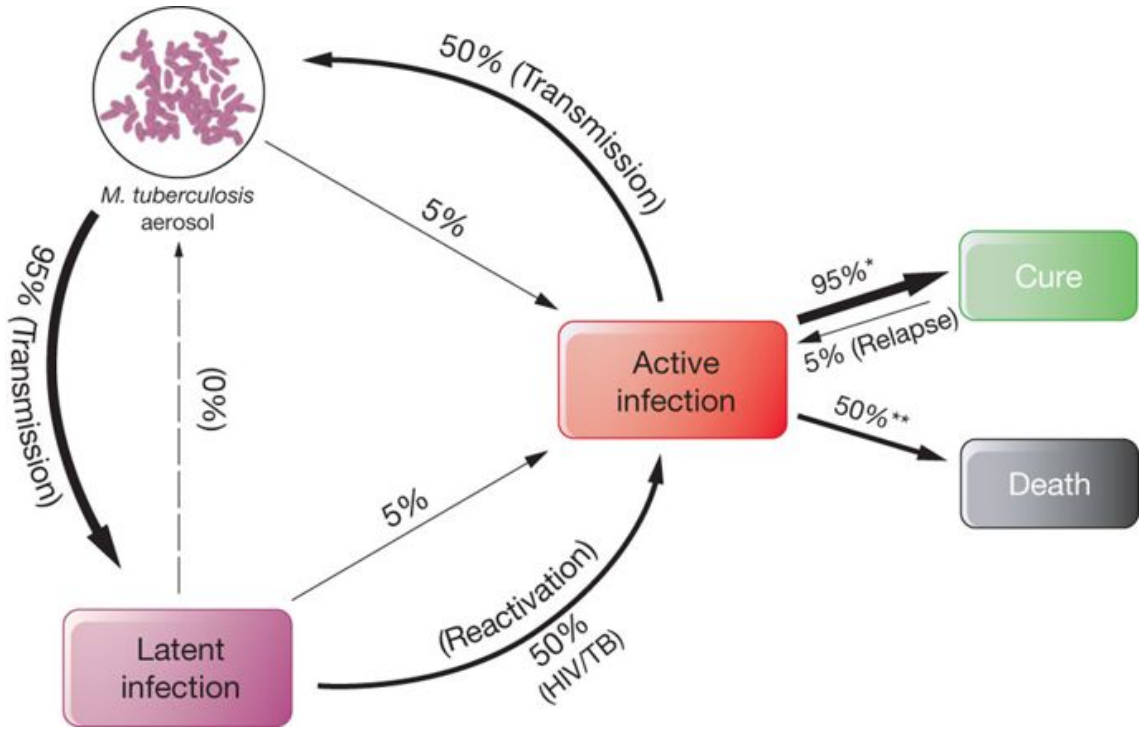


Figure 1.4: Stages of *Mtb* infection. *Mtb* transmission results in either active infection or latent infection. Latent infection can commonly develop into active infection if immunosuppressed. * refers to drug-susceptible TB, ** refers to if not treated (Koul et al., 2011).

A schematic of the life cycle of *Mtb* infection is described (Figure 1.5). Actively infected TB individuals transmit TB by the air-borne route. The bacteria, contained inside small-aerosolised droplets only 0.65-7 μm in diameter (Fennelly et al., 2004) are released by coughing or sneezing and inhaled by nearby individuals (Loudon et al., 1969). Aerosolised droplets accommodating one-three individual bacteria remain suspended in the atmosphere for several hours (Fennelly et al., 2004). Prolonged close contact with active TB infected individuals therefore is necessary to transmit the infection. Once inhaled, aerosolised droplets enter the respiratory tract, reaching the bronchi and distal alveoli where the infection establishes. Resident alveolar macrophages and other phagocytes are recruited to the site of infection and interact with the bacteria using C-type lectin receptors (Goyal et al., 2016), scavenger receptors (Zimmerli et al., 1996) and complement receptors (Hirsch et al., 1994) resulting in phagocytosis of the bacteria (Jo, 2008) (Russell, 2011). Most of the bacteria are destroyed by this innate immune response however a small number are able to survive as the bacteria prevent key defence pathways including phagosome maturation and reside inside a compartment that resembles the early endosome where they are able to replicate

intracelluarly (Ehrt et al., 2009). The macrophage also facilitates transport of the bacteria across the lung epithelium to deeper tissue (Cambier, Takaki, et al., 2014). Additional immune cells and macrophages are recruited to the infection site and form a granuloma that serves as a barrier to confine the infection from spreading however importantly this does not eradicate the infection (Chao et al., 2010). The bacteria are instead able to survive and persist in the granuloma for decades where it is called latent TB. Latent TB has a high risk of reactivation if the individual is immunocompromised e.g. receiving an organ-transplant (Aguado et al., 1997), chronic renal failure (Hussein et al., 2002) or receiving anti-TNF therapy (Solovic et al., 2010). Other lower risk factors include diabetes mellitus (Harries et al., 2011), smoking (Slama et al., 2007), steroid treatment (Jick et al., 2006) and malnutrition (Comstock, 2008). However by far the greatest risk of latent TB developing into active TB is if infected with HIV as the risk increases dramatically by 10-110 times (Ai et al., 2016). Eventually death of the granuloma infected cells causes a necrotic zone enabling bacteria to be released into the lung and further transmitted (Sharma et al., 2005).

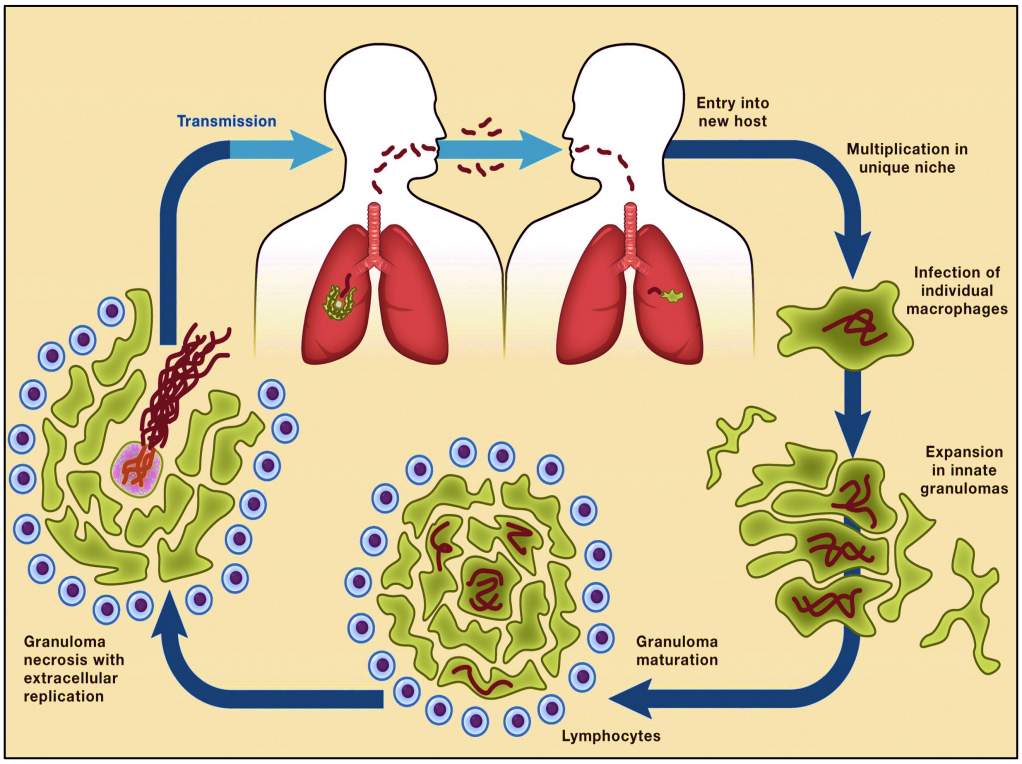


Figure 1.5: Infection life-cycle of *Mycobacterium tuberculosis* (Cambier, Falkow, et al., 2014)**.**

1.1.10 Modelling tuberculosis

As latent tuberculosis represents such an enormous disease reservoir there has been intensive research focus towards combatting the latent phase of the disease and

several models are now available to study latent TB. *In vitro* models are the most practical and widely amenable to research methods however animal infection models are better because they more accurately recapitulate the similarities in human TB.
Nutrient depravation in culture is a useful way of simulating the environment *Mtb* encounters in necrotic granulomas. For example in the Hobby and Lenert model, the carbon source from logarithmic bacterial cultures was removed by washing the cells and resuspending them into PBS. The result was bacterial growth arrest, reduced metabolism and the bacteria were refractory to isoniazid and *p*-aminosalicylic acid and showed similar antibiotic susceptibility profiles to latent TB (Hobby et al., 1957). However the best characterised model of dormancy is the Wayne and Hayes progressive hypoxia model (Wayne et al., 1996). In this model, bacteria are exposed to progressive hypoxia that mimics the low O_2 environment in necrotic granulomas. At <1 % dissolved O_2 the outer cell wall thickens and DNA replication and transcription stops. If the dissolved O_2 is reduced further to <0.06 % the bacteria are more refractory to isoniazid and rifampicin. Diluting the bacteria into O_2 rich media reversed these effects and triggered reactivation (Wayne et al., 1996). Nutrient starvation models are particularly useful to understand what nutrients *Mtb* utilises by culturing in minimal media and observing the effect of adding single substrates. As a drug discovery tool, the streptomycin dependent *Mtb* strain 18b is also a useful model because it mimics latent non-replicating bacteria in the absence of streptomycin (Benjak et al., 2016).

Animal infections models have been used to establish drug and vaccine efficacy, understand the immunological response to *Mtb* infection, the pathophysiology of disease and to model latent disease and reactivation. Mice are the most widely used models of pulmonary TB because they are practically easy to manage and have similar immunological responses to humans (A. M. Cooper, 2014). However one limitation is that they demonstrate different pathologies possibly because *Mtb* is not a natural pathogen of mice. For example one of the hallmarks of human TB are progressive caseating granulomas and lung cavitations that are not observed in mice infected with *Mtb* (Apt et al., 2009). Instead in mice most of the *Mtb* bacteria are found intracellularly in the lungs (Hoff et al., 2011). The murine, persistence model is the best characterised *in vivo* model of persistence and was developed by McCune and McDermott (R. M. McCune et al., 1966). In this model, mice were first infected with a high dose of *Mtb* and then treated with isoniazid and pyrazinamide until the recovered bacteria were no longer able to be cultured. When the antibiotics were stopped or high dose corticosteroids were added the persistent bacteria reactivated.

Guinea pigs have a very high susceptibility to *Mtb* infection and similar to in humans the disease is progressive as necrotic primary granulomas form. As Guinea pigs have a high susceptibility to disease they are used extensively for vaccine testing (U. D. Gupta et al., 2009). In contrast Rabbits are resistant to H37Rv *Mtb* however they are susceptible to *M. bovis* infection and demonstrate similar disease pathologies as humans (Manabe et al., 2003). Non-human primates are the best model of HIV/TB co-infection since they have been used for modelling simian immunodeficiency virus (SIV) infection (Lin et al., 2006). Non-human primates are evolutionary more similar to humans and therefore are used for TB drug and vaccine efficacy. However they are incredibly difficult and costly to house and there use raises significant ethical issues. Zebrafish are used extensively as an experimental animal model using the natural pathogen *M. marinum*. *M. marinum* is approximately 85 % similar to *Mtb* at the genome level (Stinear et al., 2008). *M. marinum* infection of zebrafish causes systemic disease and granuloma formation similar to in humans (Swaim et al., 2006). Zebrafish can be genetically manipulated, are easily infected and the optical properties of embryos enable fluorescent imaging therefore they are frequently used to model latent disease (Benard et al., 2012).

1.1.11 *Mtb* Survival tactics and the *Mtb* cell wall

Sputum acid-fast staining is the gold standard diagnostic for tuberculosis infection particularly in developing countries (M. Koch et al., 1965). The method is effective because of the unique and complex structure of the *Mtb* cell envelope, made up of approximately 40 % lipids by dry cell mass (Goren et al., 1974) (Minnikin et al., 1984). The high amounts of lipids make the *Mtb* cell envelope highly hydrophobic and confer an excellent barrier to impermeable antibiotics and contribute to TB being so difficult to treat (Jarlier et al., 1994). Several drugs including isoniazid (Banerjee et al., 1994), ethambutol (Telenti et al., 1997) and ethionamide (Banerjee et al., 1994) target the synthesis of the cell envelope and are used to effectively treat TB therefore the synthesis of the cell envelope has been the subject of intensive research effort (Belanger et al., 1996) (Belisle et al., 1997) (Marrakchi et al., 2000) (DeBarber et al., 2000) (Alderwick et al., 2006) (Alderwick et al., 2007).

Although the *Mtb* cell envelope is classified as Gram-positive it does not stain Gram-positive with crystal violet and its architecture more accurately resembles the Gram-negative cell envelope (Brennan et al., 1995). The current model of the *Mtb* cell envelope postulates it is made up of an inner plasma membrane surrounded by an outer membrane composed of a large mycolyl-arabinogalactan-peptidoglycan (mAGP) complex (Minnikin et al., 1984) (Figure 1.6) (Alderwick et al., 2015). The peptidoglycan

layer serves as the anchor that is bridged by arabinogalactan covalently attached to the mycolic acid layer. Mycolic acids are long chained C_{60}-C_{90} fatty acids that confer the high hydrophobicity of the cell envelope (Daffe et al., 1998). However, additional to the mAGP, several other lipids intercalate with the mycolic acid layer and plasma membrane. These include: phosphatidylinositol mannosides (PIM) (Lee et al., 1964), pthiocerol dimycocerosates (PDIM) (Daffe et al., 1998) phenolic glycolipids (PGL) (van Soolingen et al., 1997), acetyltrehalose (Takayama et al., 1977) and mannose-capped lipoarabinomannan (LAM) (Besra, Morehouse, et al., 1997). A capsule makes up the outermost layer of the cell envelope and is made up of polysaccharides, proteins and a small amount of lipids (Daffe et al., 1999). Because the cell envelope is at the interface of the bacteria and host many of its components are involved in host-pathogen interactions by mediating interactions with specialist phagocyte receptors (Stanley et al., 2013), involved in impairment of phagosome maturation in the macrophage (Russell, 2011). Inside the macrophage *Mtb* modulates the function of host cells by inhibiting maturation of phagolysosomes (Deretic et al., 2006), inhibiting acidification of phagolysosomes (Vandal et al., 2009), inhibiting apoptosis and autophagy (Briken et al., 2008) and inducing formation of the granuloma (Ehlers et al., 2012).

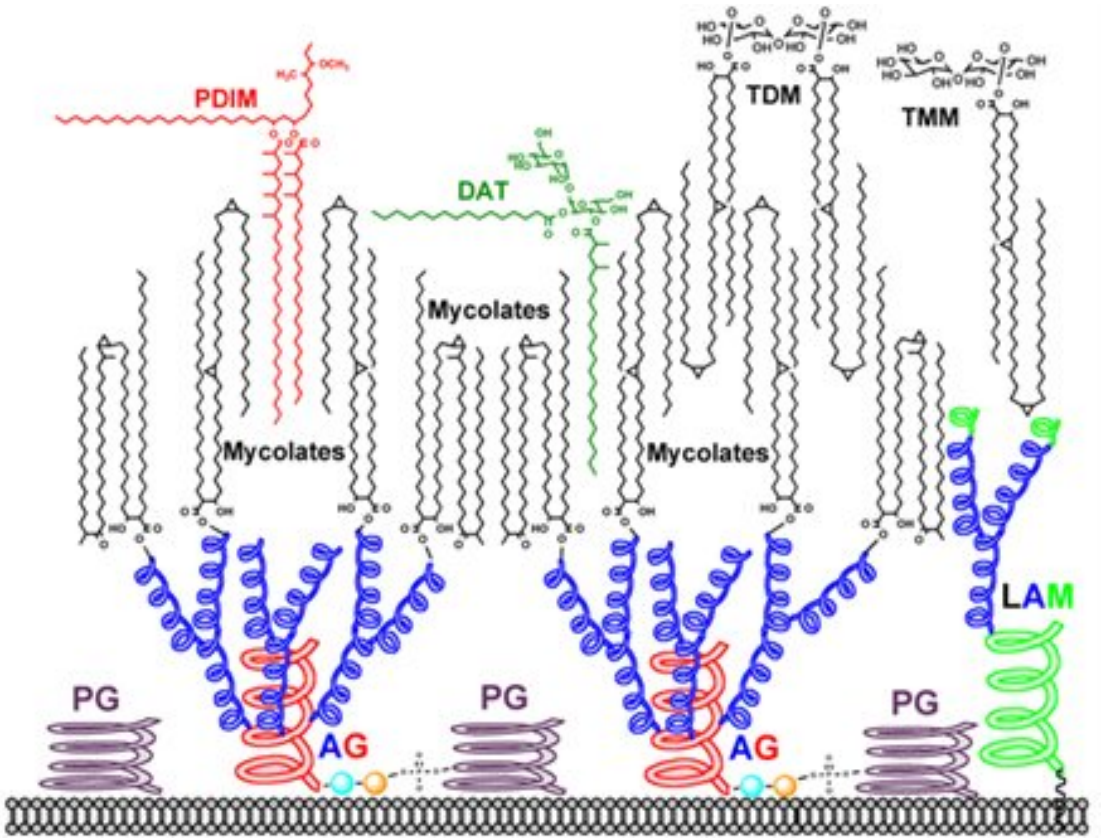


Figure 1.6: Schematic of the cell envelope of *M. tuberculosis*. Peptidoglycan (PG), arabinogalactan (AG), mycolic acids, lipoglycans: phosphatidylinositol mannosides (PIM), lipomannan (LM), lipoarabinomannan (LAM). Trehalose dimycolate (TDM), trehalose monomycolate (TMM), sulpholipids (SL), phthiocerol dimycocerosate (PDIM), di-acyl trehalose (DAT) (Alderwick, 2019).

1.1.12 Lipids found in the inner cell membrane of *Mycobacteria*

In all living organisms phospholipids are key structural components of cell membranes and make up the inner phospholipid bilayer (Figure 1.7) (Bansal-Mutalik et al., 2014). The most comprehensive quantitative analysis of the inner membrane lipids of *Mycobacteria* is from a study that successfully separated the 'outer' mycomembrane from the inner membrane of *M. smegmatis* by reverse micelle extraction (Figure 1.7) (Bansal-Mutalik et al., 2014).

Phospholipids were found to be exclusively present in the inner membrane of *M. smegmatis* and included: phosphatidylglycerol (~1.16 %), phosphatidylinositol (~0.37 %) and phosphatidylethanolamine (~0.52 %) by dry cell mass (Bansal-Mutalik et al., 2014). However by far the most abundant lipid found in the inner membrane was diacyl phosphatidylinositol dimannoside (Ac_2PIM_2) accounting for ~42 % by weight of all lipids found in the inner membrane (Bansal-Mutalik et al., 2014). Other phosphatidylinositol mannosides (PIMs) identified were either mono or diacylated and contained variable degrees of mannosylation including: $AcPIM_2$, $AcPIM_4/Ac_2PIM_4$, and $AcPIM_6/Ac_2PIM_6$

accounting for ~3% dry cell mass (Bansal-Mutalik et al., 2014). Trehalose monomycolate (TMM) and apolar lipids were also found in the inner membrane (Bansal-Mutalik et al., 2014). TMM is thought to be produced in the cytosol and transported to the cell wall to produce the mycolic acid layer (Takayama et al., 2005).

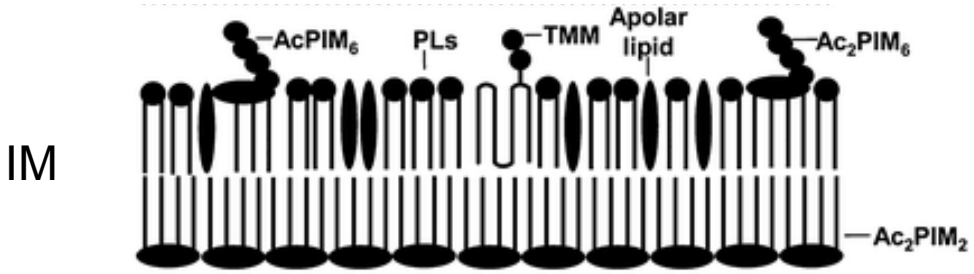


Figure 1.7: Model of the *Mycobacterial* inner membrane (Bansal-Mutalik et al., 2014)**.** IM refers to inner membrane. PL refers to the phospholipids: phosphatidylglycerol, phosphatidylinositol and phosphatidylethanolamine. $AcPIM_6$, Ac_2PIM_6 and Ac_2PIM_2 refer to phosphatidylinositol mannosides (PIM's), Ac refers to the degree of acylation and TMM refers to trehalose monomycolate.

Interestingly, phosphatidylinositol is rarely found in the membranes of prokaryotes however it is found in some actinomycetes including *Mtb* where it is a major phospholipid (Jackson et al., 2000) that serves as a lipid anchor for the cell envelope hyper-mannosylated PIM's: linear lipomannan (LM) and branched lipoarabinomannan (LAM) (Briken et al., 2004). LM and LAM are involved in virulence and have immuno-modulatory activities (Pitarque et al., 2008). Phosphatidylinositol is considered to be essential for membrane stability and cell survival and its importance is highlighted by *Mycobacteria* also encoding an inositol metabolic pathway and an essential phosphatidylinositol synthase *pgsA* gene (Jackson et al., 2000). Inositol is either scavenged or synthesised from glucose-6-phosphate and then made into phosphatidylinositol or intracellular mycothiol. Phosphatidylinositol either forms the cell membrane or is modified to form mono and diacylated PIM and the hyper-mannosylated LM and LAM (Jackson et al., 2000).

1.1.13 Lipids found in the cell membrane of macrophages

Mtb is able to survive in macrophages for decades (Pieters, 2008) however we don't fully understand exactly what nutrients it utilises. Macrophage membranes are made of several important components that *Mtb* may be able to scavenge and utilise. Macrophage membranes are made of >80 % phospholipids, the other major lipids are cholesterol, triacylglycerol (TAG) and cardiolipin (Figure 1.8) (Sahu et al., 1977). Of the different phospholipid types the major phospholipid found in the plasma membrane is phosphatidylcholine and it is mainly localised to the outer layer of the plasma

membrane (Leventis et al., 2010). Whereas the amino-group containing phospholipids: phosphatidylserine and phosphatidylethanolamine are mainly located in the inner layer of the plasma membrane (Leventis et al., 2010). Other important phospholipids of macrophages are phosphatidylglycerol, lyso-(bis)phosphatidic acid, sphingomyelin and phosphoinositides (Sahu et al., 1977). Phosphoinositides are made of phosphatidylinositol that is phosphorylated at different positions of the inositol ring. Phosphoinositides are important eukaryote intracellular signalling mediators and have critical roles during intracellular infection (Defacque et al., 2002) (De Matteis et al., 2004).

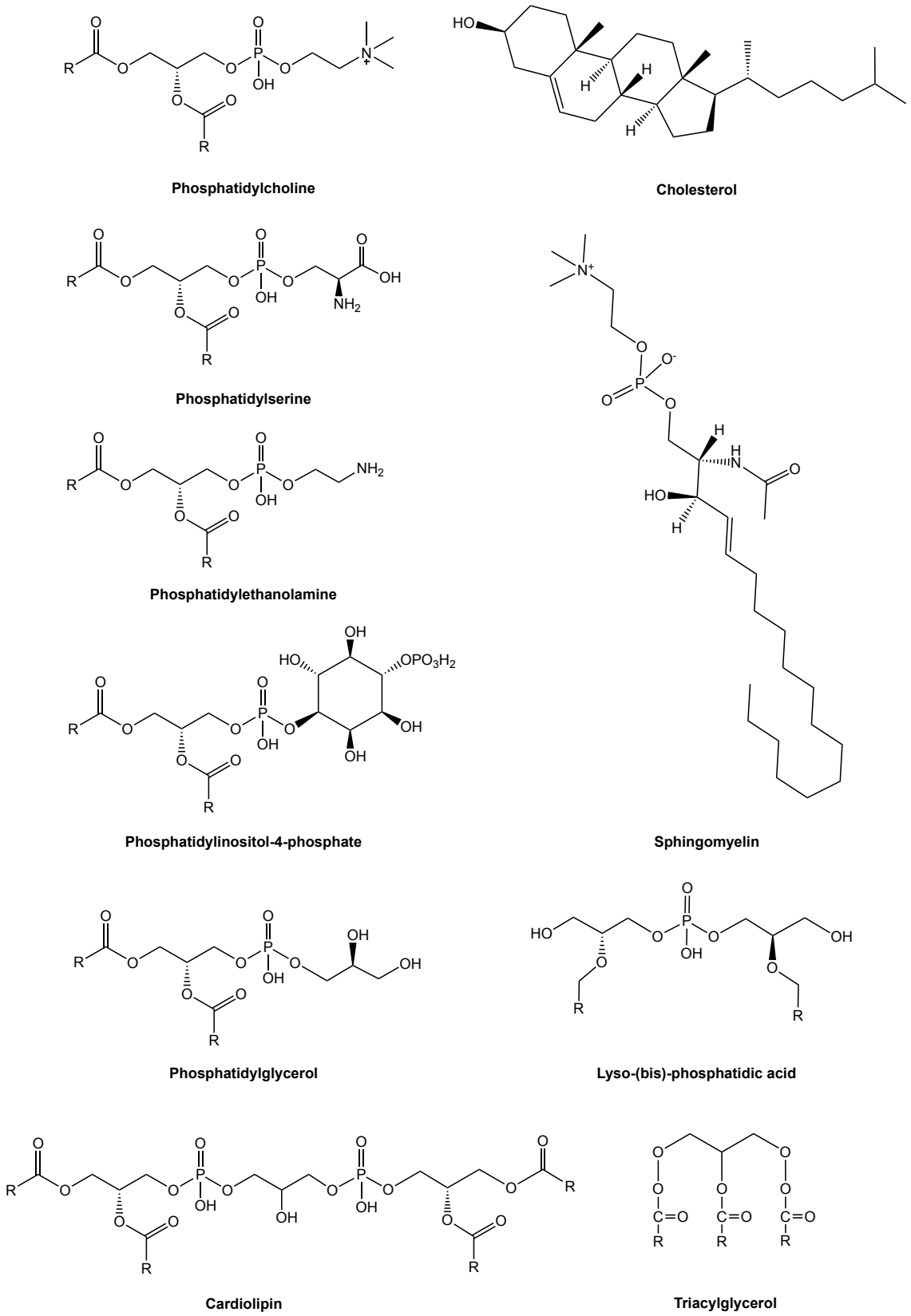


Figure 1.8: Structures of the major lipids found in macrophage membranes. R refers to the fatty acid tail. The displayed phosphatidyl-inositol-4-phosphate is a phosphoinositide however the inositol group can be phosphorylated at different positions.

One of the strategies used by *Mtb* to escape macrophage phagocytosis is by blocking phagosomal maturation through interfering with macrophage phosphatidylinositol-3-kinase signalling (Figure 1.9A) (Deretic et al., 2006). LAM in the *Mtb* cell envelope acts as a phosphatidylinositol-3-phosphate analogue and inhibits macrophage cytosolic Ca^{2+} release leading to blockade of the phosphatidylinositol-3-kinase cascade. As a result phagosomes do not mature because phosphatidylinositol-3-phosphate is not produced on liposome and phagosome membranes (Vergne et al., 2003). However the macrophage is also able to utilise its membrane phospholipids for innate defences (Callahan et al., 2000). Redistribution of macrophage phosphatidylserine triggers efferocytosis, an apoptotic cell engulfment process where un-infected macrophages phagocytose the apoptotic *Mtb* infected macrophage to kill and restrict *Mtb* growth (Figure 1.9B) (Callahan et al., 2000).

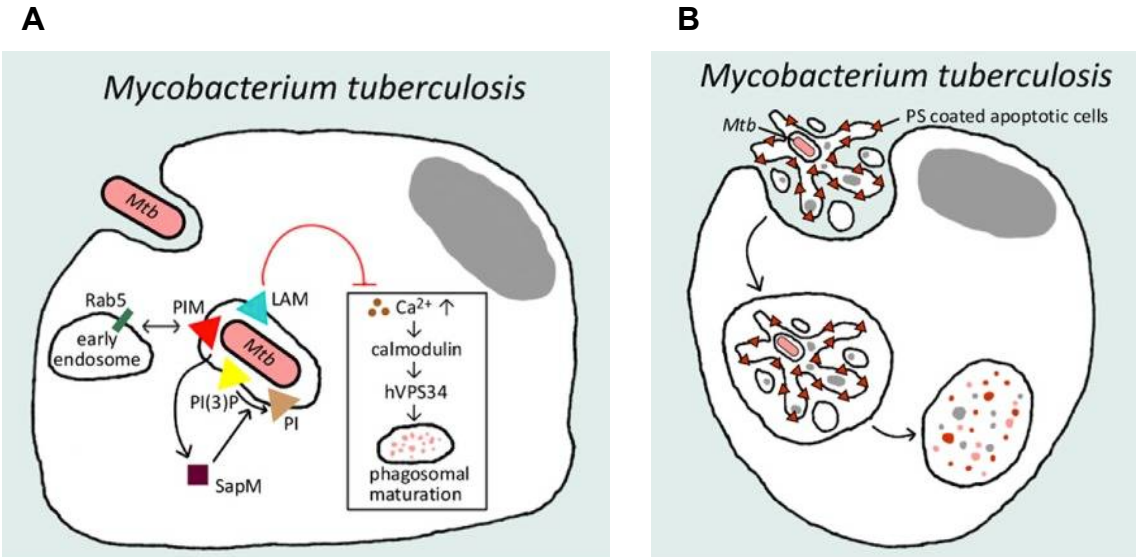


Figure 1.9: Manipulation of host phospholipids by *Mtb* and macrophage phospholipid defence strategy (Teng et al., 2017)**.** A) *Mtb* escapes macrophage phagocytosis by blocking phagosomal maturation using LAM to interfere with macrophage phosphatidylinositol-3-kinase signalling. B) Redistribution of macrophage phosphatidylserine leads to efferocytosis, phagocytosis of apoptotic *Mtb* infected macrophages.

1.1.14 *Mycobacterium tuberculosis* nutrient acquisition

Mtb resides in a nutrient limited environment and the complete nutritional requirements required for its survival are not well characterised. By identifying the specific nutrients it utilises, targeted methods could be developed to restrict nutrient uptake and effectively starve the bacteria. However to date, the complete nutritional profile utilised by *Mtb* during infection in the host is not known (Titgemeyer et al., 2007). Studies in the early 20th Century (Edson, 1951) investigated the growth of *Mtb* in minimal media and different carbon sources and found that *Mtb* most effectively grows on glycerol or sodium salts of fatty acids. In 1925 Novy and Soule (Novy, 1925) developed respiratory

exchange to determine the oxygen consumption and carbon dioxide production of *Mtb* grown on nutrient agar containing either glycerol or glucose. Later Loebel, Shorr and Richardson (Loebel, 1930) studied *Mtb* respiration in suspended cultures containing phosphate buffer and single substrates and found that either glycerol or sodium salts of fatty acids were the most effective stimulants of respiration. Whereas glucose produced only a weak stimulation and arabinose, fructose, mannose and inositol produced no significant effects (Loebel, 1930). A more extensive study by Nakamura (Nakamura, 1938) measured oxygen consumption of *Mtb* in the presence of 25 different substrates and found that glycerol was the only substrate that produced an increase with a small stimulatory effect for glutamate. However it wasn't until the 1950's when 7H9 broth and 7H10 agar media were developed and found to be instrumental to routinely culture *Mtb* (Middlebrook et al., 1954) (Middlebrook et al., 1958). These media effectively support the growth of *Mtb* if supplemented with nutrients such as glycerol, oleic acid, albumin and dextrose provided by Middlebrook OADC/ADC enrichment. The primary carbon sources in these media are glycerol, dextrose and the fatty acid, oleic acid. It was later found that *Mtb* is also able to utilise tween 80 as a carbon source that is frequently added to cultures to prevent cell clumping (Lyon et al., 1963). It is presumed that due to the slow growth rate of *Mtb*, doubling every 24 hours, glycerol is able to freely diffuse across the membrane at a sufficient rate to support cell growth (Paula et al., 1996). However it is argued that all molecules are imported by specific transporter proteins (Kell, 2015). Intracellular glycerol kinases may convert glycerol to glycerol-3-phosphate where it enters central carbon metabolism pathways to generate ATP (Niederweis, 2008). However the conditions that *Mtb* encounters during infection are far more complex compared to in culture.

Studies have found that *Mtb* utilises carbohydrates during the early stages of infection (Sassetti et al., 2003). However it is thought that the host immune response triggers a switch to fatty acids becoming the primary nutrient source (Kalscheuer et al., 2010). Evidence for fatty acid utilisation is supported by the glyoxylate cycle being essential for *Mtb* survival in mice as isocitrate lyases, malate synthases and β-oxidation enzymes that convert fatty acids to acetate that flow into the citric acid cycle are required to sustain growth and persistence in macrophages (McKinney et al., 2000) (Schnappinger et al., 2003) (Dubnau et al., 2005) (Timm et al., 2003). However the types of fatty acids that may be utilised, where they are sourced and how they are imported are not fully understood.

Cholesterol is an important molecule found in eukaryote cell membranes including macrophages and studies have demonstrated that *Mtb* utilises cholesterol and that it is required for growth and persistence in animal infection models (J. C. Chang et al., 2009) (Van der Geize et al., 2007). *Mtb* encodes >80 genes (Van der Geize et al., 2007) dedicated to cholesterol import and metabolism as it provides the precursors, propionyl-CoA and methyl malonyl-CoA of important cell wall lipids including PDIM and SL as well as pyruvate that enters the citric acid cycle (Wilburn et al., 2018). Cholesterol has also been found to accumulate in the granuloma in animal infection models (Kim et al., 2010).

One of the major challenges that *Mtb* faces is how to import these different nutrients across the outer and inner impermeable cell membranes, similar to how it is so difficult for antibiotics to penetrate the cell envelope. To date we don't fully understand what transporters *Mtb* utilises to facilitate nutrient import. Recently, sophisticated NMR (Somashekar et al., 2011) and mass spectrometry, metabolomics methods (Nandakumar et al., 2015) have investigated the metabolites that fluctuate during *Mtb* infection of macrophages and in animal models. High-resolution magic angle spinning NMR spectroscopy revealed that in *Mtb* infected guinea pig granuloma, the rate of metabolic flux for carbohydrates, amino acids, phospholipid components and other metabolites increases (Somashekar et al., 2011). This suggested that *Mtb* potentially utilises alternative energy sources through the glycolytic, pentose phosphate and citric acid cycle pathways in developing granuloma (Somashekar et al., 2011). ^{13}C metabolomics studies have also revealed metabolic pathway remodelling. For example, during hypoxia the *Mtb* citric acid cycle slows and increases production of succinate in order to flexibly sustain membrane potential, ATP synthesis and anaplerosis (Eoh et al., 2013). However further investigations are required because *Mtb* has probably evolved multiple complex strategies to utilise nutrients in the varied and scarce environments it encounters in the macrophage, the granuloma, the extracellular environment and specific *Mtb* susceptible tissues.

1.1.15 Carbohydrate transporters of *Mycobacterium tuberculosis*

Bioinformatics analysis of the *Mtb* H37Rv and *M. smegmatis* genomes revealed that *Mtb* only encodes 5 putative carbohydrate transporters whereas *M. smegmatis* the non-pathogenic environmental saprophyte encodes an enormous 28 putative carbohydrate transporters (Figure 1.10), suggesting it is well adapted to import environmental carbohydrates (Titgemeyer et al., 2007). Therefore the limited number of putative carbohydrate transporters encoded by *Mtb* may reflect the nutrient scarce host cell macrophage environment (Titgemeyer et al., 2007).

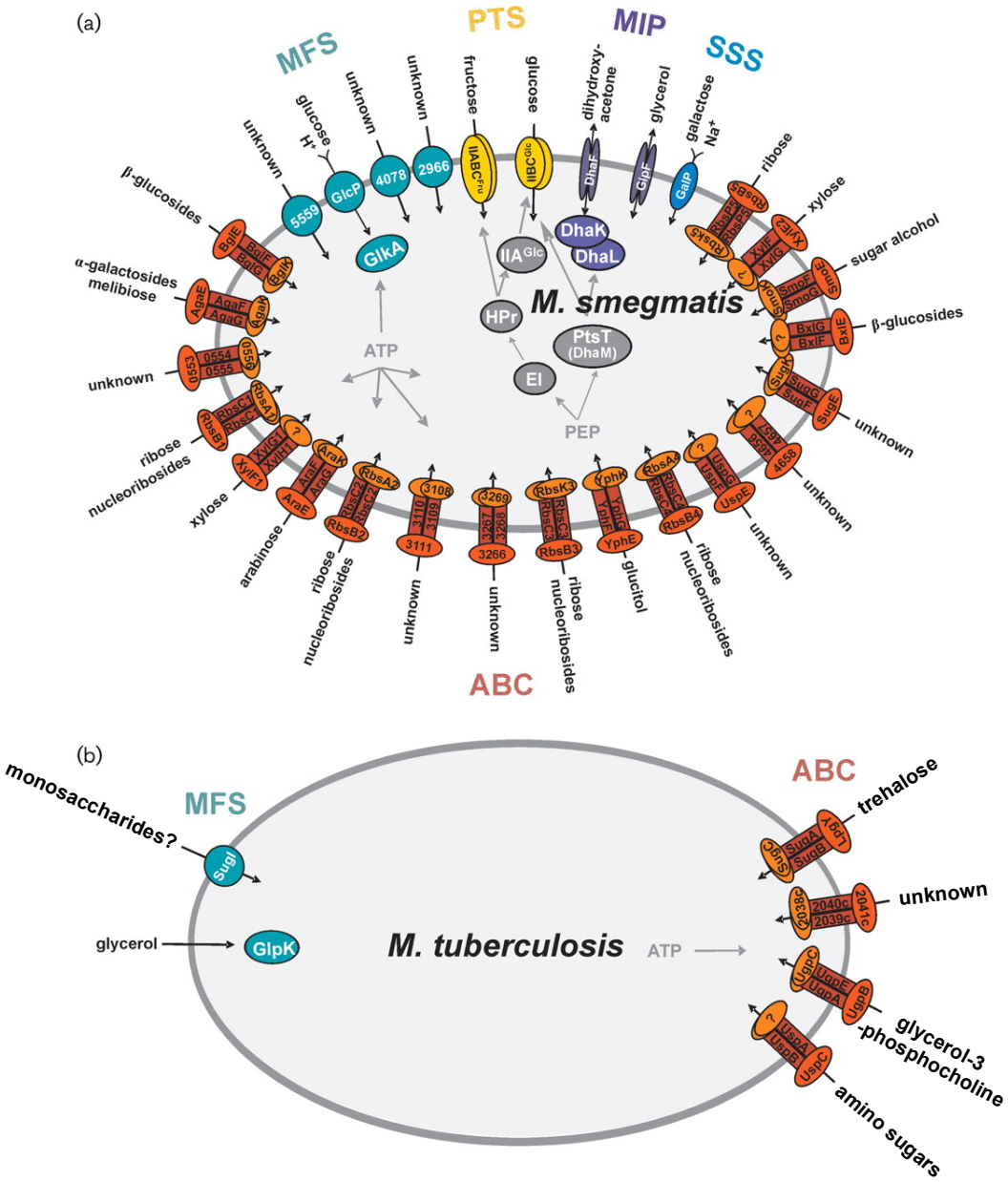


Figure 1.10: Comparison of the putative carbohydrate transporters of *Mycobacterium smegmatis* and *Mycobacterium tuberculosis* (Niederweis, 2008). MFS refers to major facilitator superfamily, PTS refers to phosphotransferase system, MIP refers to major intrinsic protein family, SSS refers to sodium solute superfamily and ABC refers to ATP-binding cassette transporter.

Of the 5 putative transporters encoded by *Mtb*, 4 belong to the ATP-binding cassette (ABC) transporter class and only 1 is a permease that belongs to the major facilitator superfamily (Braibant et al., 2000). ABC transporters facilitate nutrient import in an ATP hydrolysis dependent alternating access mechanism (Locher, 2016). They are typically composed of 4 subunits (Figure 1.11), the substrate binding protein (SBP) specifically binds to cognate ligand in the periplasm, the membrane spanning domains (MSD's) and the nucleotide binding proteins (NBP's) that hydrolyse ATP to ADP and P_i (Locher,

2016). In contrast, permeases of the major facilitator superfamily are single integral membrane proteins made up of 12-14 transmembrane α-helices and also facilitate nutrient import by alternating access (Pao et al., 1998). The predicted *Mtb* carbohydrate permease of the major facilitator superfamily is SugI and the 4 ABC carbohydrate transporters include: LpqY-SugABC, UspABC, Rv2041c-2040c-2039c-2038c and UgpABCE (Figure 1.10) (Niederweis, 2008). Little knowledge about the substrate preference can be predicted because the sequence similarities of these transporters are very low outside of the *Mycobacteria* genus (Titgemeyer et al., 2007). Many of these transporters are encoded by other *Mycobacteria* pathogens as well as non-pathogenic species. Overall SugI, UspABC, Rv2041c-2040c-2039c-2038c and UgpABCE are poorly characterised. LpqY-SugABC was predicted to bind to maltose however it was found to actually bind to trehalose (Kalscheuer et al., 2010). Further studies of these transporters could reveal important nutrient uptake pathways utilised by *Mtb* during infection and identify new targets for drug development.

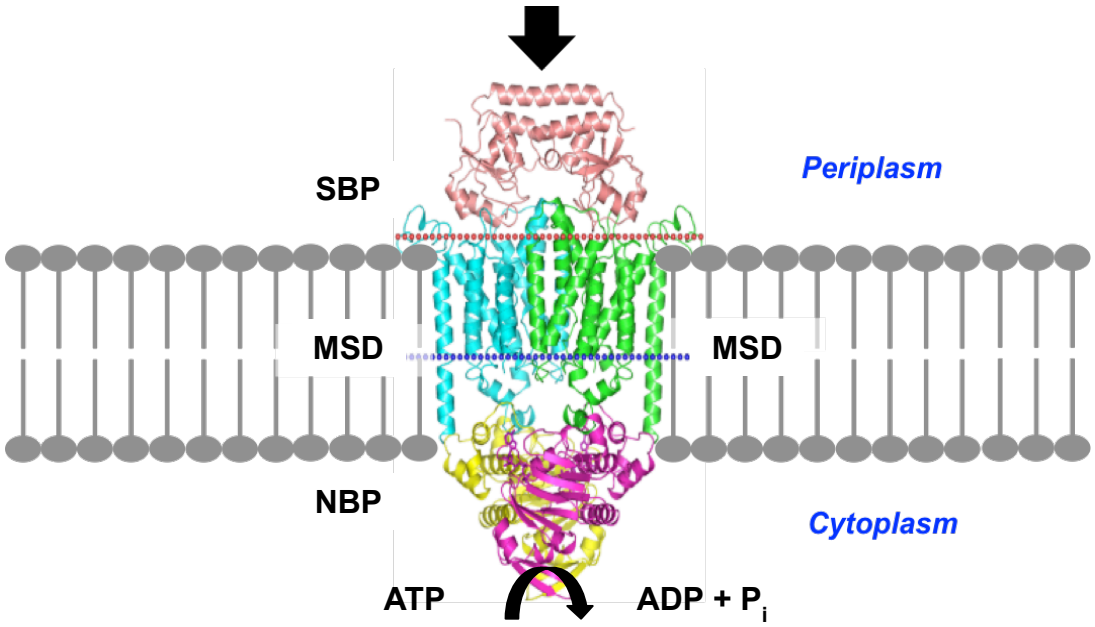


Figure 1.11: Architecture of an ATP-binding cassette (ABC) importer. SBP refers to the substrate binding protein, MSD refers to the membrane spanning domains, and NBP refers to the nucleotide binding proteins. The location of the subunits relative to the cytoplasm and periplasm are indicated.

1.1.15.1 *Mtb* SugI

SugI is a predicted integral membrane protein and member of the major facilitator superfamily. SugI is predicted to be composed of 12 transmembrane α-helices (ExPASy TMPred) (K. Hofmann, 1993). By amino acid sequence homology SugI is most similar to the glucose permease GlcP (28 %) of *S. coelicolor* and to the galactose and arabinose transporters, GalP and AraE (24 %) of *E. coli*. As all of these similar

transporters import monosaccharides SugI may also be involved in the transport of monosaccharides (Niederweis, 2008). Recently a study identified a mutation in SugI that generated a premature stop codon and caused N-terminal truncation of SugI. This mutation in SugI was associated with resistance to the anti-TB antibiotic, D-cycloserine (Chen et al., 2017). Interestingly D-cycloserine has a similar overall structure to the furanose pentose monosaccharides therefore SugI may be involved in the import of arabinose and assembly of the arabinogalactan layer of the cell envelope (Figure 1.12). However to date SugI has not been biochemically characterised. As it is a single transmembrane protein, SugI is anticipated to be difficult to express (Schlegel et al., 2014).

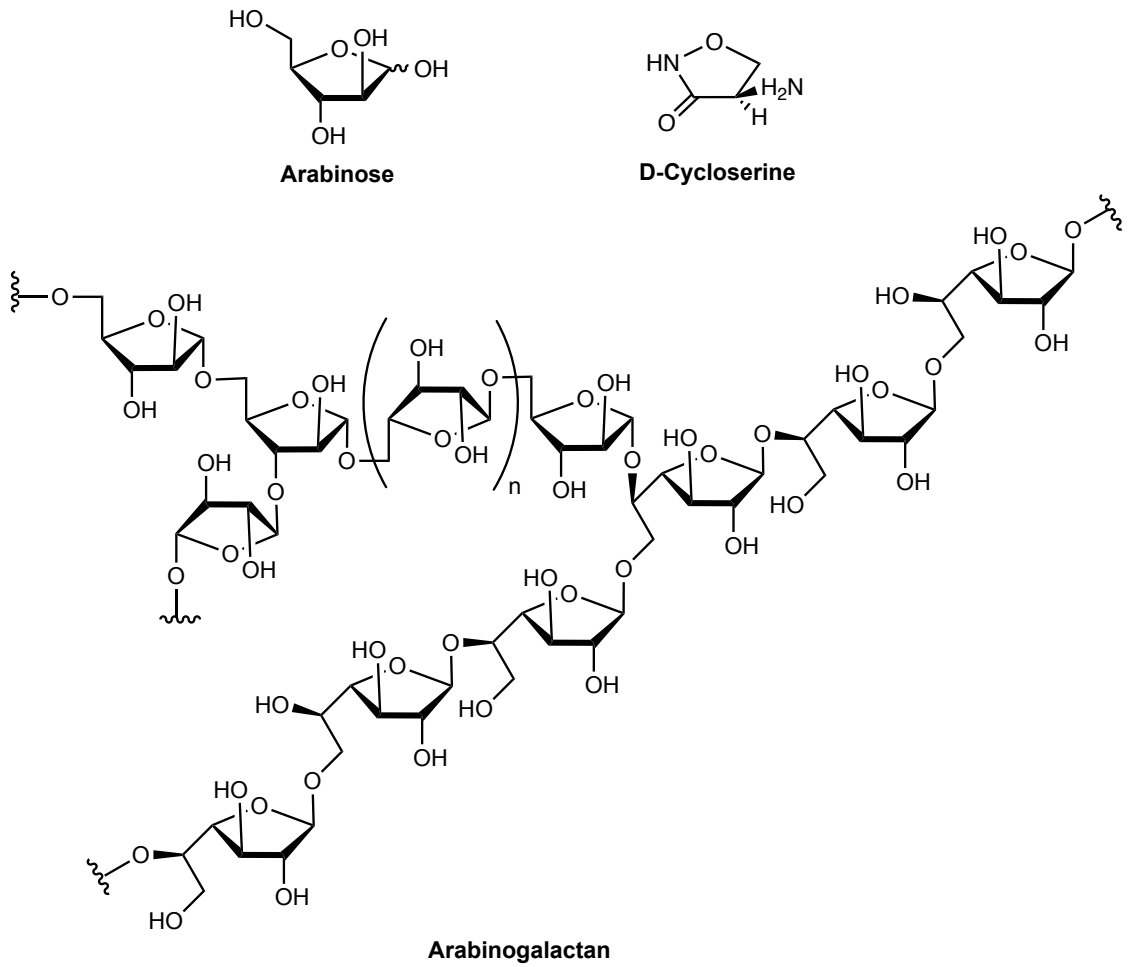


Figure 1.12: Comparison of the structures of the furanose pentose monosaccharide, arabinose, D-Cycloserine and *Mtb* arabinogalactan.

1.1.15.2 *Mtb* LpqY-SugABC

LpqY-SugABC is especially interesting because it was the first of the 5 predicted carbohydrate importers of *Mtb* to be identified as critical for virulence (Kalscheuer et al., 2010). LpqY-SugABC is an ABC transporter and is specific for the uptake of the

disaccharide trehalose. Trehalose is a sugar not present in mammals as only a small amount of dietary trehalose is ingested and enzymatically digested by trehalase enzymes to be utilised as glucose (Gudmand-Hoyer et al., 1996). However, in *Mtb* trehalose is released as a by-product of mycolic acid synthesis by the activity of the Ag85 complex that synthesises the lipid trehalose dimycolate (TDM) (Takayama et al., 2005) (Ronning et al., 2000). The released free trehalose is recycled and transported by LpqY-SugABC where it is used to synthesise trehalose monomycolate (TMM). MmpL3 exports TMM across the inner membrane where it is processed by Ag85 to TDM and forms the mycomembrane (Figure 1.13) (Wolber et al., 2017). Interestingly a mutant *Mtb* LpqY-SugABC strain showed strong attenuation in an immunocompromised mice model and a severe growth defect in the mice lungs during the acute phase of infection however it could persist during the chronic infection and demonstrates a role for this transporter in the early stages of infection (Kalscheuer et al., 2010). In a separate study trehalose was chemically modified to have azido, deoxy and fluoro-sterochemical modifications at positions 2-, 5- and 6- and when added to *M. smegmatis* cells inhibited growth and the ability for biofilm formation (Wolber et al., 2017). Although studies of LpqY-SugABC in whole *Mtb* cells have been carried out it has not yet been biochemically characterised to bind to trehalose.

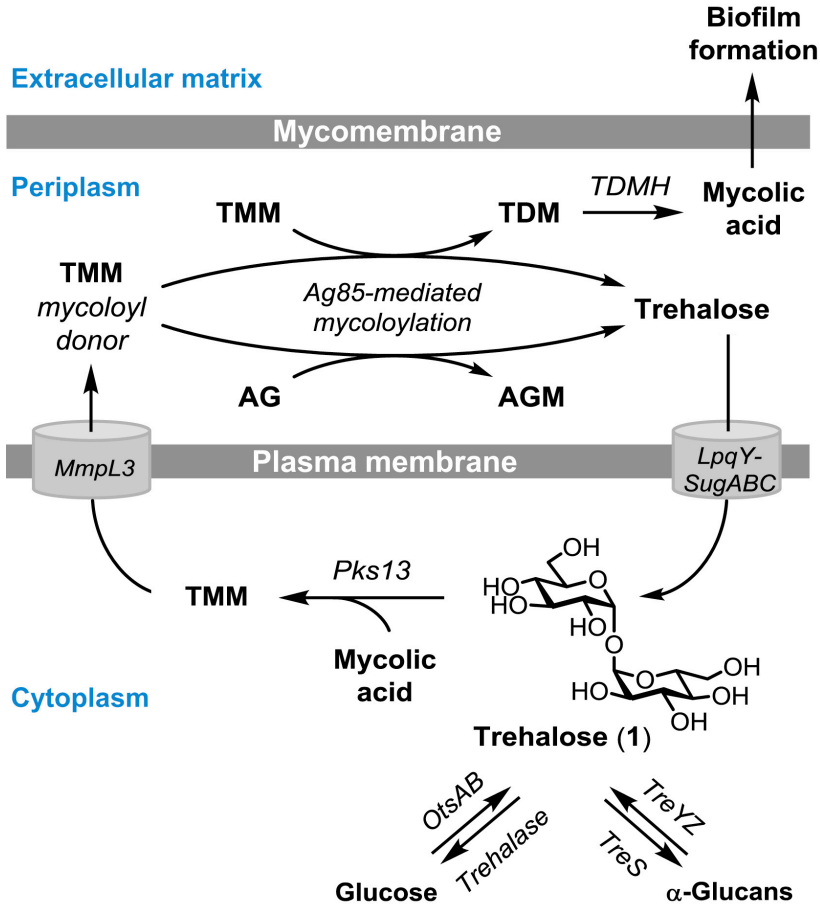


Figure 1.13: Potential trehalose recycling pathway. LpqY-SugABC imports free trehalose that is either converted to α-Glucans, glucose or trehalose monomycolate (TMM). TMM is exported by MmpL3 and converted to TDM by Ag85. TDM is converted to mycolic acids and free trehalose is released and recycled (Wolber et al., 2017).

1.1.15.3 *Mtb* UspABC

UspABC is interesting because it may be involved in the uptake of cell wall peptidoglycan fragments (Fullam et al., 2016). UspABC is also an ABC transporter and similar to LpqY-SugABC it is also found in *Mycobacterium leprae* indicating that it may be critical for the core life cycle of *Mycobacteria* (Vissa et al., 2001). Recently the structure of the substrate binding protein, UspC was determined by X-ray crystallography in the open ligand unbound conformation (Figure 1.14A) and revealed it had the general architecture of substrate binding proteins. UspC was also found to bind to amino-sugars specifically with an amino group at position C2 or C3. Binding was detected to D-glucosamine-6-phosphate and chitobiose, carbohydrates that structurally resemble peptidoglycan fragments (Figure 1.14B). This implicates a potential role for UspABC in peptidoglycan recycling (Fullam et al., 2016). Also as UspC is present near the inner membrane it is located in the perisplamic space proximal to the cell wall peptidoglycan of which fragments could be released by the activity of resuscitation promoting factors that have muramidase and amidase activities

(Kana et al., 2008). Further studies are underway to characterise the precise substrate of the UspABC transporter.

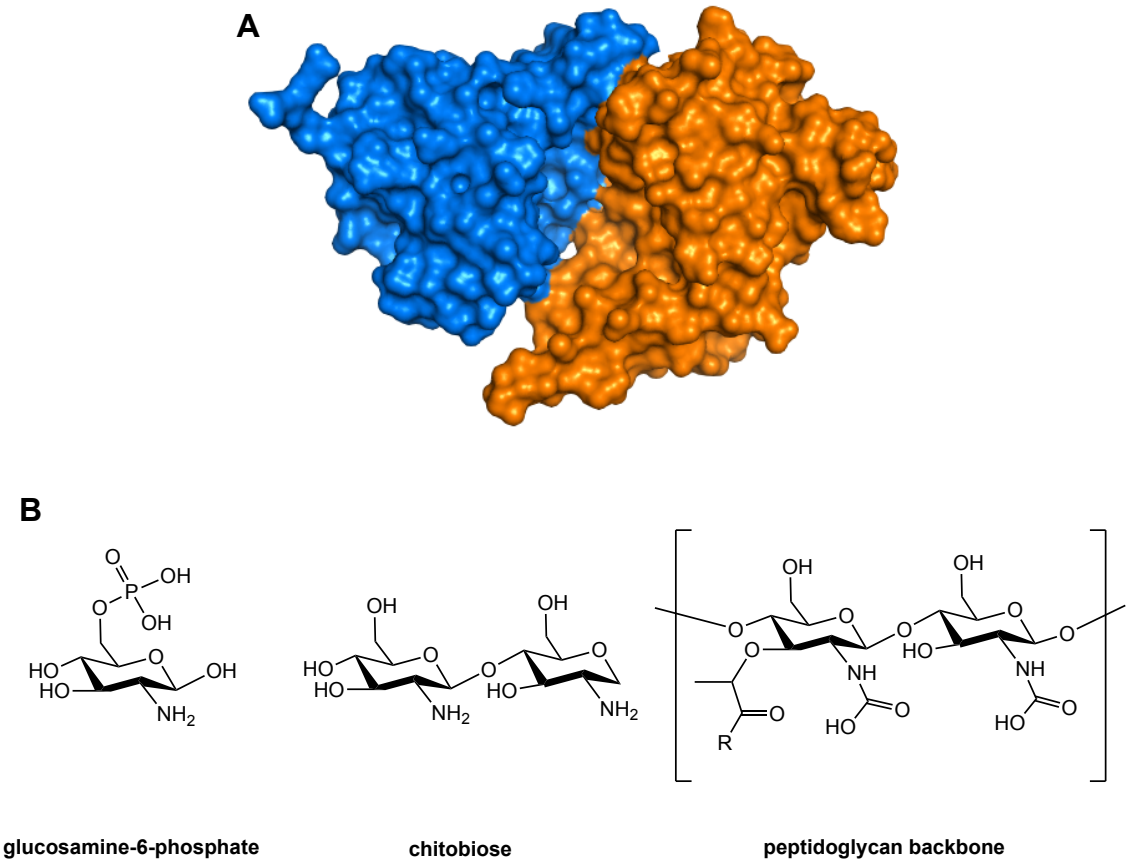


Figure 1.14: Overall structure of *Mtb* UspC and ligands identified to bind (Fullam et al., 2016)**.** A) Overall structure of *Mtb* UspC. B) Structures of glucosamine-6-phosphate, chitobiose and the peptidoglycan backbone.

1.1.15.4 *Mtb* Rv2041c-2040c-2039c-2038c

Rv2041c-2040c-2039c-2038c is a predicted ABC transporter and may be important for when *Mtb* adapts to stressful conditions (Shin et al., 2009). Interestingly, Rv2041c was found to be highly expressed in *Mtb* at low pH 6.0 and hypoxia (13 % O_2 in *Mtb* infected macrophages) (Shin et al., 2009). These conditions are similar to the intracellular phagosome environment. Rv2041c is predicted to be a substrate binding protein however its substrate is currently not known. A previous study expressed Rv2041c however Rv2041c was insoluble in inclusion bodies and was solubilised with guanidine hydrochloride. Immunoblotting of TB patient and mice infected with *Mtb* sera identified the presence of antibodies against Rv2041c therefore It could be used to potentially increase the sensitivity of serodiagnosis (Shin et al., 2009).

1.1.15.5 *Mtb* UgpABCE

UgpABCE is the final predicted *Mtb* carbohydrate ABC transporter and is the focus of this PhD project. The UgpABCE transporter is predicted to have the typical architecture of *Mtb* ATP-binding cassette importers (Locher, 2009) (Figure 1.15). The structure of the ligand unbound substrate-binding protein UgpB was previously determined by X-ray crystallography (PDB ID 4MFI) (Figure 1.15) (Jiang et al., 2014). In these previous studies, UgpB was identified to bind to glycerol-3-phosphocholine (GPC) the head-group of phosphatidylcholine, which is one of the most abundant lipids found in eukaryote cell membranes (Yang et al., 2018). However phosphatidylcholine is not an abundant phospholipid in the *Mycobacteria* membranes (Bansal-Mutalik et al., 2014). The identification of binding of UgpB to GPC by ITC suggests that UgpABCE may be involved in the uptake of host derived catabolised lipids where GPC is imported and further metabolised to be used as a phosphate or carbon nutrient supply or enter phospholipid biosynthesis. Importantly we do not know the structural basis of how GPC binds to UgpB or indeed if it is the physiological ligand and other alternative higher affinity ligands may still exist. The *ugpB* gene is not essential from *in vitro* transposon hybridisation studies (DeJesus et al., 2017) however UgpB is up regulated *in vivo* during *Mtb* infection of guinea pigs (Kruh et al., 2010). Intriguingly, the vaccine strain *Mycobacterium bovis* BCG also encodes a UgpABCE transporter however it has a frame shift mutation in UgpB making it non-functional (Brosch et al., 2007). Serial passage of *M. bovis* onto potato slices containing glycerol is thought to have led to the loss of virulence of the resulting *M. bovis* BCG strain (Calmette, 1922) and may have also selected for this UgpB mutation because in contrast the pathogenic species, *Mycobacterium bovis* has an intact UgpB gene (Brosch et al., 2007).

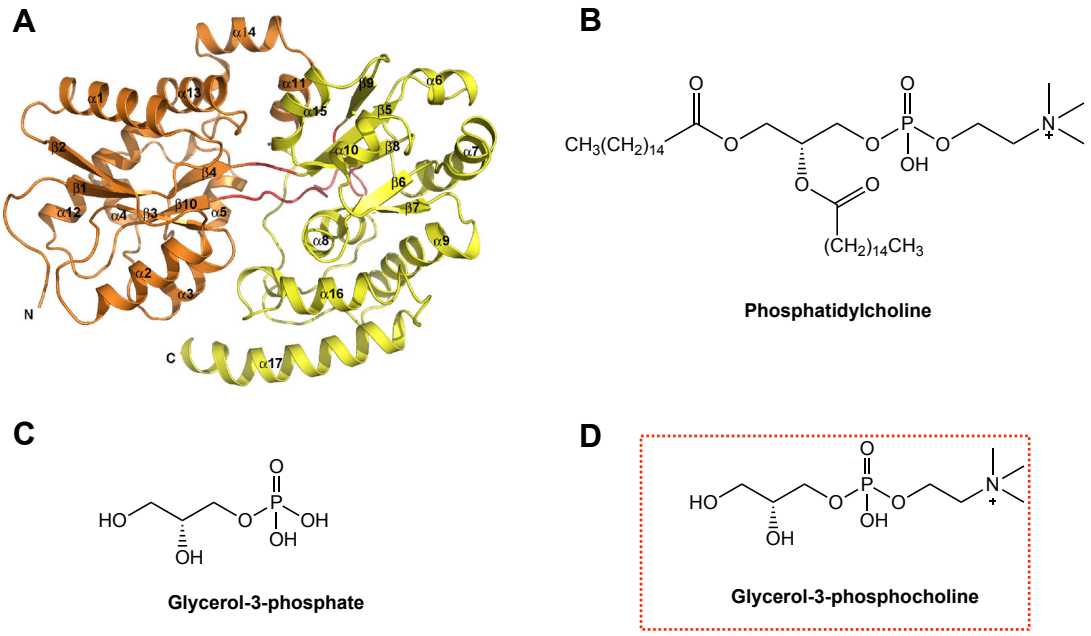


Figure 1.15: Overall structure of *Mtb* UgpB and ligands (Jiang et al., 2014)**.** A) Overall structure of *Mtb* UgpB (PDB ID 4MFI). B) Structure of phosphatidylcholine the phospholipid derivative of glycerol-3-phosphocholine. C) Structure of glycerol-3-phosphate that did not show binding to *Mtb* UgpB. C) Structure of glycerol-3-phosphocholine that showed binding to *Mtb* UgpB.

1.1.16 Glycerol-3-phosphate/Glycerophosphodiester uptake in other bacteria

Mtb may encounter phosphate depletion inside the phagosome (Rengarajan et al., 2005) therefore it has likely evolved several different strategies to acquire this essential nutrient. *Mtb* encodes two putative inorganic phosphate uptake systems (Pst) however it is not fully known how these systems are regulated (Tischler et al., 2016). In contrast the *Mtb* UgpABCE transporter is potentially adapted to specifically uptake organic phosphate. Many different species of bacteria also encode UgpABCE transporters for the uptake of glycerol-3-phosphate (G3P) in phosphate-limited conditions (Luo et al., 2009) (Wuttge et al., 2012) (Chandravanshi et al., 2016).

E. coli encodes a UgpABCE ABC-transporter however compared to *Mtb* UgpB the substrate binding protein *E. coli* UgpB binds to both G3P and GPC (Wuttge et al., 2012). The structure of G3P bound *E. coli* UgpB was solved by X-ray crystallography (PDB ID 4AQ4) (Wuttge et al., 2012). However at the amino acid level *Mtb* UgpB and *E. coli* UgpB are only 26.4 % similar over 83 % read (BLAST) and suggests they may be functionally different. In addition to UgpABCE, *E. coli* also encodes a GlpT proton symporter that actively transport G3P into the cell indicating that there are multiple systems available for G3P import (Lemieux et al., 2004). Interestingly, from comparative genomics *Mtb* does not encode a GlpT proton symporter and therefore may have adapted the UgpABCE transporter specifically for G3P or GPC import.

A wide variety of marine environmental bacteria are also predicted to encode a UgpABCE transporters which they may use to import dissolved organic phosphorous as dissolved organic phosphorous is an essential marine nutrient (Luo et al., 2009). The largest phyla predicted to encode UgpABCE transporters are the α-proteobacteria. However the study found that the mean number of UgpABCE transporter genes was greater in samples collected from the open ocean compared to coastal waters suggesting that the UgpABCE transporter is important for uptake of glycerol-3-phosphate from the oceans (Luo et al., 2009).

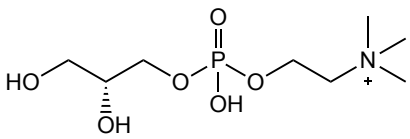
Extremophile bacteria are also predicted to encode UgpABCE transporters. The Gram-negative thermophile *Thermus thermophillus* encodes a predicted UgpABCE transporter that based on computational analysis also binds to GPC and has a similar ligand-binding pocket to *E. coli* UgpB (Chandravanshi et al., 2016).

Interestingly other human pathogens encode a predicted UgpABCE transporter including *Mycoplasma pneumonia* the causative agent of atypical pneumonia (Schmidl et al., 2011). However *M. pneumoniae* only encodes the following genes: *ugpC*, *ugpA* and *ugpE* (MPN134, MPN135, MPN136) and does not encode a predicted *ugpB* gene for the substrate binding protein. Interestingly *M. pneumonia* thrives on lung epithelia as a source of phospholipids for nutrients (E. J. Veldhuizen et al., 2000). The UgpACE transporter of *M. pneumoniae* may actually have a different substrate specificity because *M. pneumonia* is not able to grow on the single carbon source glycerol-3-phosphate (Schmidl et al., 2011). In addition to a UgpACE transporter, *M. pneumonia* also encodes GlpU a permease of the major facilitator superfamily that is responsible for import of glycerol-3-phosphocholine (GPC) (Grosshennig et al., 2013). Interestingly as *Mtb* does not encode a GlpU permease instead it may have adapted its UgpABCE transporter to specifically import GPC. Eukaryotes also possess transporters for glycerophosphodiesters for example, the yeast *Saccharomyces cerevisiae* encodes GIT1 a permease used to import glycerophosphoinositol during inositol restricted conditions in order to synthesise the essential phosphatidylinositol membrane lipids (Patton-Vogt et al., 1998).

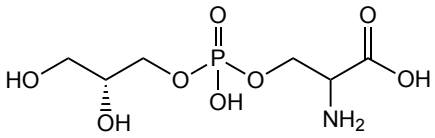
Collectively, species comparisons do not provide much additional information about the function and substrate specificity of the *Mtb* UgpABCE transporter therefore a more thorough biochemical analysis of *Mtb* UgpB was the focus of this PhD project.

1.1.17 Are glycerophosphodiesters the physiological substrate of *Mtb* UgpABCE

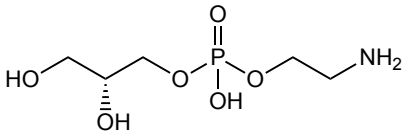
Phospholipid recycling by *Mycobacteria* is an emerging area of research accelerated by the search to understand what nutrients *Mtb* utilises during infection. During infection *Mtb* occupies many different host niches however it resides primarily inside macrophages (Pieters, 2008). Macrophage membranes are made of the major phospholipids of which many glycerophosphodiesters are derived including: phosphatidylcholine, phosphatidylserine, phosphatidylethanolamine, phosphatidylglycerol and phosphoinositides as described previously (1.1.13) (Figure 1.8). Phosphatidylcholine, specifically the diacyl dipalmitoylphosphatidylcholine is also an abundant extracellular component of lung surfactant (Veldhuizen et al., 2000). In *Mtb* infected guinea pig granuloma, GPC is increased with a concomitant decrease in the level of phosphatidylcholine suggesting that phospholipids may be catabolised by the *Mtb* bacilli (Somashekar et al., 2011). Previous studies of *Mtb* UgpB described earlier (1.1.20) indicated that *Mtb* UgpB binds to the glycerophosphodiester, glycerol-3-phosphocholine (GPC) derived from the phospholipid phosphatidylcholine. However to date it hasn't been investigated if the *Mtb* UgpABCE transporter is able to bind and import other lipids with alternative phospholipid head-groups that are potentially derived from host macrophage membranes (Figure 1.16).



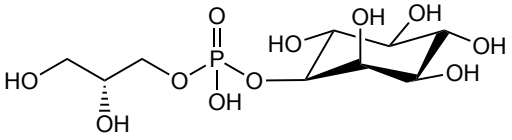
Glycerol-3-phosphocholine (GPC)



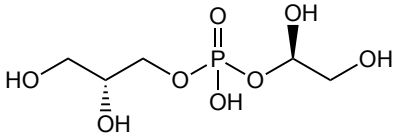
Glycerophosphoserine (GPS)



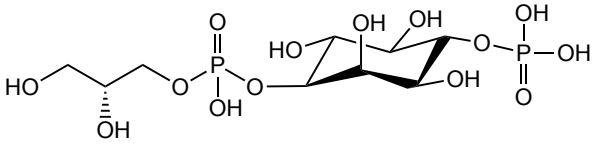
Glycerophosphoethanolamine (GPE)



Glycerophosphoinositol (GPI)



Glycerophosphoglycerol (GPG)



Glycerophosphoinositides

Figure 1.16: Structures of glycerophosphodiesters that may be derived from macrophage cell membranes.

1.1.18 Phospholipase enzymes

Phospholipase enzymes are required to convert phospholipids to glycerophosphodiesters (Dennis, 2015). It is not yet established if *Mtb* is able to utilise glycerophosphodiesters derived from macrophage membranes or lung surfactant however both the host and *Mtb* do encode genes that have phospholipase activity. Additionally, host phospholipases are well established to have critical roles in eukaryote phospholipid biosynthesis and cellular signalling and may be involved in generating substrates (Burke et al., 2009).

Phospholipase enzymes are classified as having activities that catalyse the hydrolysis of phospholipids (Figure 1.17). These activities result in either the release of hydrophobic fatty acid tails or soluble components of the glycerophosphate head-group (Figure 1.17) (Dennis, 2015). Phospholipase A_1 hydrolyses at position *sn*-1 to release a fatty acid tail and phospholipase A_2 hydrolyses at position *sn*-2 to release the other fatty acid tail. Phospholipase C hydrolyses the glycerophosphate bond to release a phosphor head-group. Phospholipase D hydrolyses the phosphodiester bond to release phosphatidic acid and a free head-group. Phospholipase C enzymes are encoded by *Mtb* and found to be up regulated in phosphate starvation conditions (Le Chevalier et al., 2015). Recently a phospholipase C *Mtb* deletion mutant was identified to not be able to grow if phosphatidylcholine was the only phosphate source indicating that the ability to degrade phospholipids is important for *Mtb* bacilli growth (Le Chevalier et al., 2015). Further investigations are being carried out to establish the importance of *Mtb* phospholipase C's. However it is the activities of phospholipase A_1 and A_2 that are required to produce glycerophosphodiesters, which have also been identified in *Mtb* (1.1.19).

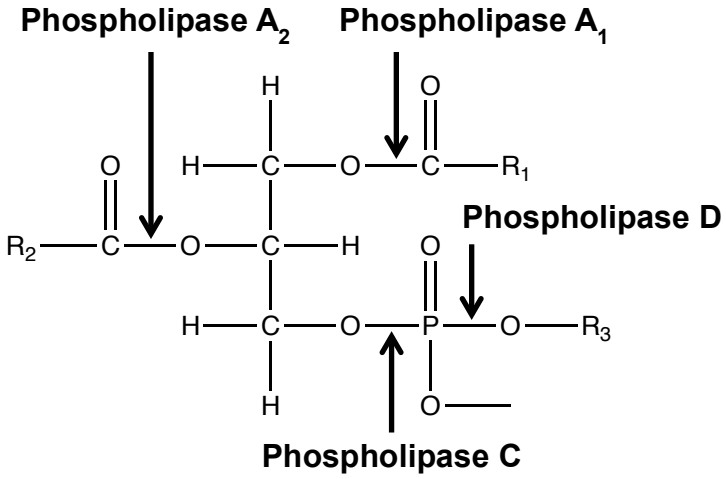


Figure 1.17: Different phospholipase activities and the phospholipid positions cleaved.

1.1.19 Phospholipase A_1 and A_2 activities produce glycerophosphodiesters

In order to produce glycerophosphodiesters, phospholipase A_1 and A_2 enzymatic activity is required. Phospholipase A activities have been identified in *Mtb*. For example two secreted cutinases Rv1984 and Rv3452 were found to hydrolyse medium chain carboxylic esters and monoacylglycerols (Brust et al., 2011) Rv3542 also has phospholipase A_2 activity (Schue et al., 2010). Another enzyme annotated as Rv3802 was found to hydrolyse the following host derived and *Mycobacteria* derived phospholipids: phosphatidylcholine, phosphatidylethanolamine and phosphatidylserine (Goins et al., 2018). However, secreted host macrophage phospholipases A_1 and A_2 enzymes may also be involved in producing glycerophosphodiesters (B. E. Schneider et al., 2014). For example lysosomal phospholipase A_2 is implicated in the ability to control intracellular *Mycobacteria* growth however the exact lipid mediator is not known (Teng et al., 2017). Establishing if *Mtb* UgpABCE binds to a range of glycerophosphodiesters would provide a firm basis to rigorously investigate the source of the substrates.

1.1.20 Cleavage of glycerophosphodiesters either extracellular or intracellular

Instead of directly importing glycerophosphodiesters some organisms secrete enzymes that degrade glycerophosphodiesters to glycerol-3-phosphate. For example *E. coli* secretes a Ca^{2+}-dependent GlpQ enzyme into the periplasm that has been identified to hydrolyse the following glycerophosphodiesters (Tommassen et al., 1991): glycerophosphoethanolamine (GPE), glycerophosphocholine (GPC), glycerophosphoglycerol (GPG) and bis(glycerophospho)glycerol to glycerol-3-phosphate. Glycerol-3-phosphate can then be imported by the proton symporter GlpT (Figure 1.18) (Larson et al., 1983) (Huang et al., 2003).

In addition to the periplasmic GlpQ, *E. coli* also encodes a cytosolic glycerophosphodiesterase UgpQ adjacent to the UgpABCE transporter in its genome (Ohshima et al., 2008). UgpQ converts imported glycerophosphodiesters to glycerol-3-phosphate (Figure 1.18). *Mtb* also encodes two genes annotated to have glycerophosphodiesterase activity: *glpQ1* and *glpQ2* that may be involved in converting imported glycerophosphodiesters to glycerol-3-phosphate that enter glycerophospholipid metabolism or glycolysis. However these *Mtb* GlpQ enzymes in contrast to *E. coli* GlpQ are found in the cytosol and therefore likely to be active intracellularly (Larrouy-Maumus et al., 2013). *Mycoplasma pneumonia* also encodes a GlpQ enzyme immediately downstream of the permease GlpU (Grosshennig et al., 2013) that specifically cleaves GPC to G3P, and is potentially used as an energy source as *M. pneumonia* thrives in lung epithelia (Veldhuizen et al., 1998).

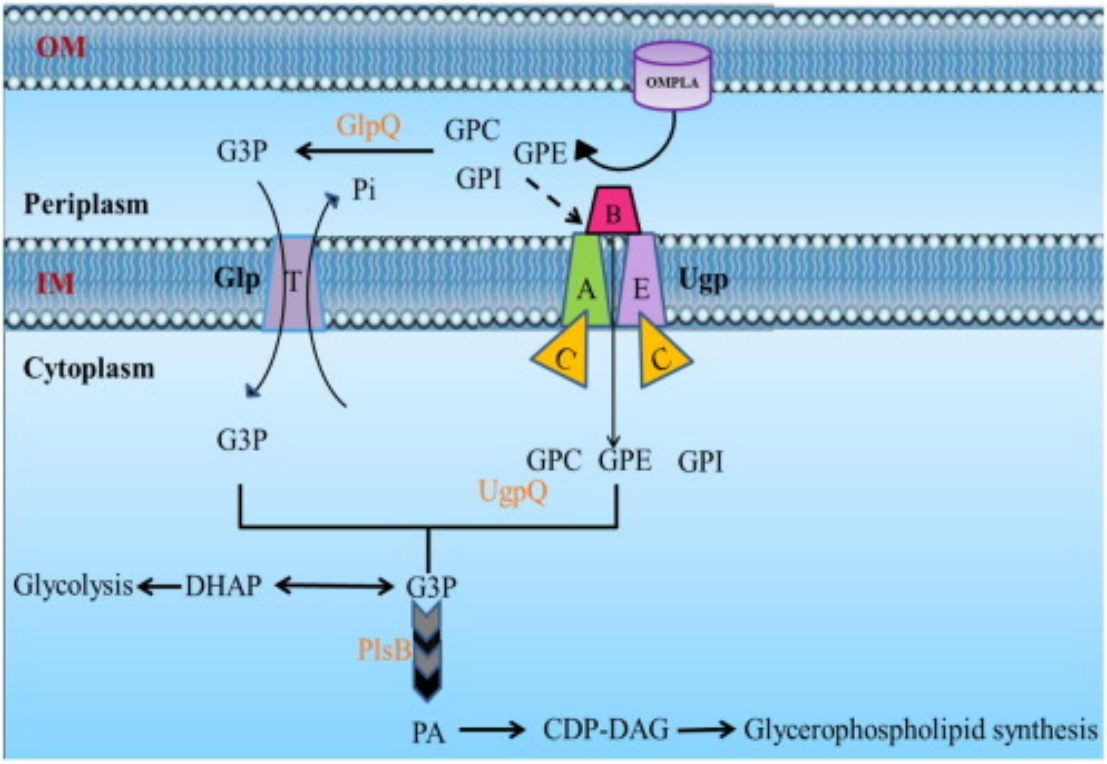


Figure 1.18: GlpT permease and UgpABCE transporter of *E. coli* (Chandravanshi et al., 2016). *E. coli* secretes GlpQ that converts glycerophosphodiesters to glycerol-3-phosphate. Glycerol-3-phosphate is transported by GlpT. UgpABCE is also able to bind to glycerophosphodiesters and glycerol-3-phosphate (Wuttge et al., 2012).

Possibly because of the different host environment encountered by *Mtb* compared to *E. coli*, the *Mtb* UgpABCE transporter may have evolved to specifically recognise GPC directly instead of relying on a secreted glycerophosphodiesterase enzyme to convert extracellular GPC to G3P. This is supported by a previous study that indicated *Mtb* UgpB only binds to GPC and not to G3P (Jiang et al., 2014). However GPC may not be the only glycerophosphodiester recognised by the *Mtb* UgpABCE transporter and we wanted to investigate if UgpB was promiscuous for any alternative lipid head-groups that it may scavenge throughout its infection life cycle.

1.1.21 A potential *Mtb* glycerophospholipid recycling pathway

Interestingly a putative phospholipid-recycling pathway was recently identified in *Mtb* (Larrouy-Maumus et al., 2013). Rv1692 encodes a glycerol-3-phosphate phosphatase that was proposed to be the final enzyme in a glycerophospholipid-recycling pathway that converts glycerol-3-phosphate to free phosphate and glycerol (Figure 1.19). Glycerol may then be utilised directly by the bacteria as a nutrient source. Rv1692 was not essential for the *in vitro* survival of *Mtb* however deletion of Rv1692 led to increases in the amount of glycerol phosphate, phosphatidylglycerol and glycerophosphoinositol suggesting it is part of a phospholipid recycling pathway

(Larrouy-Maumus et al., 2013). A large increase in radiolabelled ^{14}C glycerophosphoethanolamine and glycerophosphoinositol was also observed in the Rv1692 deletion mutant suggesting that degradation of these polar lipid head groups was decreased in the absence of Rv1692.

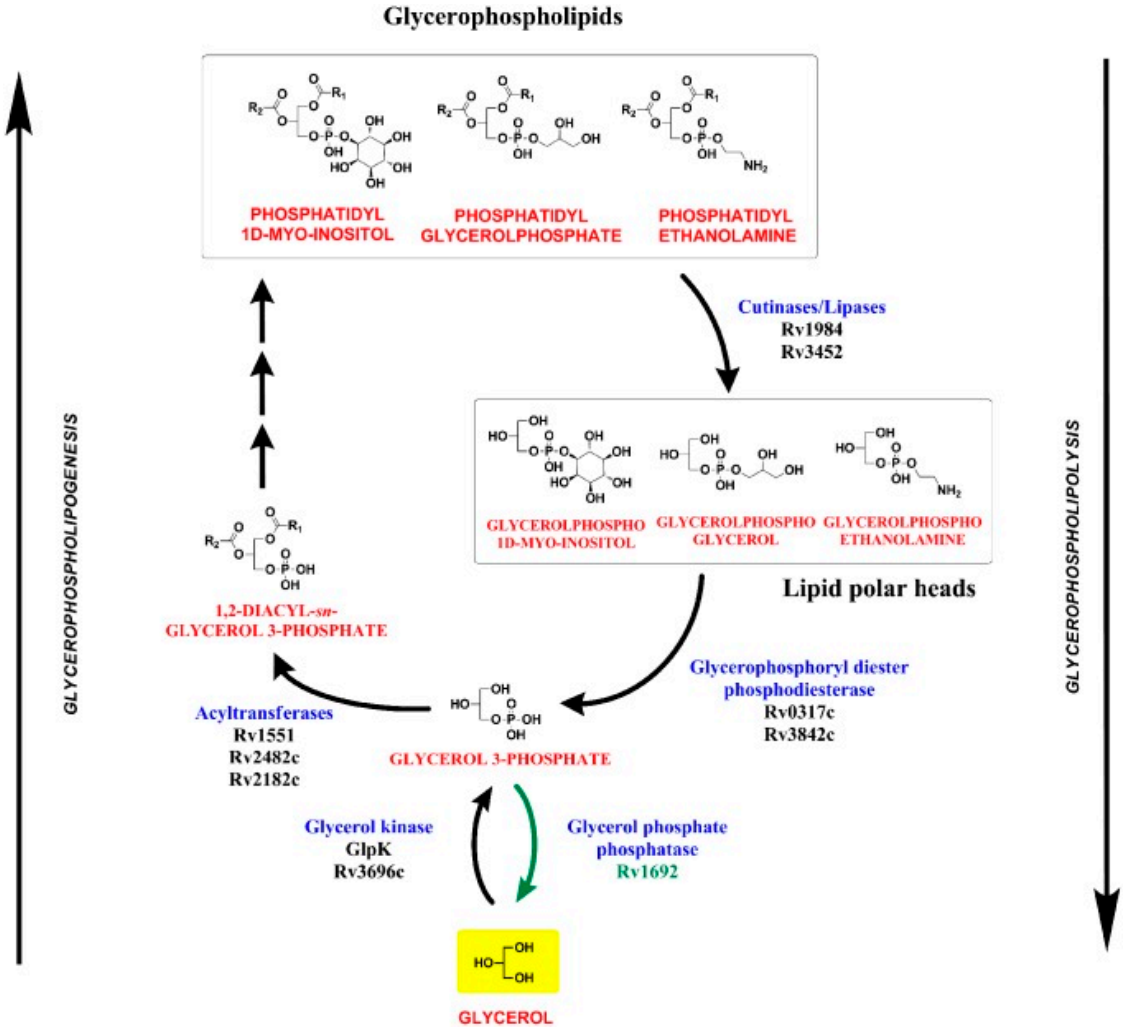


Figure 1.19: Proposed glycerophosphodiester catabolism pathway in *Mycobacterium tuberculosis* (Larrouy-Maumus et al., 2013). Glycerophospholipids are degraded to lipid polar heads by cutinases/lipases. Lipid polar heads are degraded to glycerol-3-phosphate by glycerophosphoryl diester phosphodiesterases. Glycerol-3-phosphate is converted to glycerol by glycerol phosphate phosphatase and glycerol enters metabolic pathways. Alternatively glycerol-3-phosphate is converted to glycerophospholipids.

Many of the genes in this putative glycerophospholipid-recycling pathway need to be fully characterised. For example two genes previously described called Rv1984 and Rv3452 have been shown to have lipase and esterase activity (Brust et al., 2011) and may be important for producing glycerophosphodiesters however host phospholipases and previously uncharacterised enzymes may also be involved.

1.1.22 Summary

To conclude, the *Mtb* UgpABCE ABC-transporter may have adapted to bind to glycerophosphodiesters. It is not yet known if binding is limited to glycerol-3-phosphocholine, or if glycerophosphodiesters containing alternative head-groups could also bind. Due to the recent identification of putative phospholipid recycling pathways, the *Mtb* UgpABCE transporter may have an important role in host phospholipid catabolism to survive during nutrient scarcity.

1.2 Hypotheses and Aims

The overall hypothesis of this study was that *Mtb* uses the UgpABCE transporter to transport a wide range of glycerophosphodiester ligands. The promiscuity of this transporter may provide carbon and phosphate nutrient sources that are further metabolised by the bacteria. This may aid the ability for *Mtb* to survive in nutrient scarce environments such as that found inside macrophages. The details of how I tested these hypotheses are outlined in the hypotheses and aims sections of chapters 3-6.

1.3 Specific aims

The specific aims of the work presented in this study are:

- Chapter 3: Express and purify *Mtb* UgpB and genetic variants.
- Chapter 4: Biochemical characterisation of *Mtb* UgpB.
- Chapter 5: Structural analysis of *Mtb* UgpB ligand interactions.
- Chapter 6: Express and purify *Mtb* UgpC and determine if it is a functional ATPase. Express and purify *Mtb* UgpAE.

Chapter 2: Materials and methods

2.1 Materials

2.1.1 Materials

All chemicals and reagents used for biochemical studies were purchased from Carbosynth, Fisher Scientific, New England Biolabs, Sigma Aldrich, Tebu Bio or Thermo Scientific unless otherwise stated. All kits were used according to the manufacturers instructions unless otherwise stated. The University of Warwick Life Sciences media preparation service provided routine microbiological media and antibiotics.

Table 2.1: Buffers

Buffer	Components
Tris acetic acid EDTA buffer (TAE)	10 mM tris-HCl 0.1 mM ethylenediaminetetraacetic acid (EDTA) pH 8.0
Phosphate buffered saline (PBS)	137 mM NaCl 2.7 mM KCl 10 mM phosphate pH 7.4
Phosphate buffered saline tween-20 (PBST)	137 mM NaCl 10 mM NaH_2PO_4 2.5 mM KCl 2.5 mM KH_2PO_4 0.05 % (v/v) tween-20 pH 7.4
Tris buffered saline	20 mM tris-HCl 0.9% (w/v) NaCl pH 7.4
Tris buffered saline tween-20 (TBST)	20 mM tris-HCl 0.9% (w/v) NaCl 0.1 % (v/v) tween-20 pH 7.4
UgpB lysis buffer	20 mM NaH_2PO_4 500 mM NaCl 0.1 % (v/v) triton X-100 pH 7.4

UgpB buffer 1	20 mM NaH_2PO_4 500 mM NaCl pH 7.4
UgpB wash buffer	20 mM NaH_2PO_4 1 M NaCl pH 7.4
UgpB dialysis buffer 1	25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) 150 mM NaCl pH 7.0
UgpB dialysis buffer 2	25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) 150 mM NaCl 10 % (v/v) glycerol pH 7.0
UgpC buffer	20 mM tris-HCl 300 mM NaCl 20 % glycerol 5 mM ATP pH 7.5
pWaldo purification buffer	20 mM tris-HCl 300 mM NaCl 10 % glycerol pH 7.5
10X SDS-PAGE gel running buffer	24.8 mM tris-HCl 0.192 M glycine 0.1 % (w/v) sodium dodecyl sulphate (SDS)
2X SDS loading dye	100 mM tris-HCl 4 % (w/v) sodium dodecyl sulphate (SDS) 0.2 % (w/v) bromophenol blue 20 % (v/v) glycerol 200 mM dithiothreiotol (DTT) pH 6.8
Gel wash buffer	50 mM ammonium bicarbonate 50 % (v/v) ethanol
Reducing buffer	10 mM dithiothreitol (DTT) 50 mM ammonium bicarbonate
Alkylation buffer	55 mM iodoacetamide 50 mM ammonium bicarbonate

Trypsin buffer	2.5 ng/μL trypsin solution 50 mM ammonium bicarbonate
Low salt phosphate buffer	25 mM NaH_2PO_4 100 mM NaCl 10 % (v/v) glycerol pH 7.0
Malachite green phosphate assay buffer	20 mM tris-HCl 300 mM NaCl 4 % glycerol 1 mM $MgCl_2$ pH 7.5
Western blot transfer buffer	25 mM tris-HCl 192 mM glycine 10 % MetOH
Bio-Layer buffer A	25 mM NaOAc 150 mM NaCl 10% glycerol pH 5.5
Bio-Layer buffer B	25 mM NaH_2PO_4 150 mM NaCl 10 % glycerol pH 7.0
Bio-Layer buffer C	25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) 150 mM NaCl 10 % glycerol pH 7.0

Table 2.2: Cloning primers

Name	Use	Sequence (5'-3')
UgpB_T_pYUB_F	Constructed by Dr James Harrison and used to clone truncated *Mtb* UgpB (amino acid residues 1-34) pYUB1062, *NdeI* restriction site	aaaaaa*catatg*ggttccggcccaatcgacttctgg
UgpB_T_pYUB_R	Constructed by Dr James Harrison and used to clone truncated *Mtb* UgpB pYUB1062, *HindIII* restriction site	aaaaaa*aagctt*gccatgccccgccagcttccg
UgpC pET SUMO_F	Clone *Mtb* UgpC pET SUMO, *BamHI* restriction site	aaaa*ggatcc*atggctaacgttc
UgpC pET SUMO_R	Clone *Mtb* UgpC pET SUMO, *HindIII* restriction site	aaaa*aagctt*tcagcgaagccgg
UgpC pYUB1062_F	Clone *Mtb* UgpC pYUB 1062, *NdeI* restriction site	aaaa*catatg*gctaacgttcagt
UgpC pYUB1062_R	Clone *Mtb* UgpC pYUB 1062, *HindIII* restriction site	aaaa*aagctt*gcgaagccgggtc
UgpC pYUB1062 GFP_R	Clone *Mtb* UgpC pYUB1062 GFP, *HindIII* restriction site	aaaa*aagctt*tgcgaagccgggt
UgpAE pWaldo_F	Clone *Mtb* UgpAE pWaldo, *NdeI* restriction site	aaaaaa*catatg*gcggcgccgca
UgpAE pWaldo_R	Clone *Mtb* UgpAE pWaldo, *BamHI* restriction site	ttttttg*gatcc*tcccttgaccg

Sequences in italics refer to the restriction endonuclease recognition sites *NdeI*, *HindIII* or *BamHI*.

Table 2.3: Sequencing primers

T7_F	Sequence pYUB1062 and pWaldo constructs	taatacgactcactataggg
T7_R	Sequence pYUB1062 and pWaldo constructs	ccctatagtgagtcgtatta
SUMO_F	Sequence UgpC pET SUMO construct	agattcttgtacgacggtattag
pWaldo-R	Sequence pYUB1062-GFP and pWaldo constructs	ttcaccctctccactgacag

Table 2.4: *Mtb* UgpB site directed mutagenesis primers

Tyr78Ala_F	Mutate *Mtb* UgpB residue Tyr78Ala	ggcaaggac**gcc**gacgaggtg
Tyr78Ala_R	Mutate *Mtb* UgpB residue Tyr78Ala	cacctcgtc**ggc**gtccttgcc
Asp102Ala_F	Mutate *Mtb* UgpB residue Asp102Ala	cgttttgctcgac**gcc**cgatggtggttcc
Asp102Ala_R	Mutate *Mtb* UgpB residue Asp102Ala	ggaaccaccatcg**ggc**gtcgagcaaaacg
Ser153Ala_F	Mutate *Mtb* UgpB residue Ser153Ala	ccgtatgctcgc**gcg**acgccgctgttc
Ser153Ala_R	Mutate *Mtb* UgpB residue Ser153Ala	gaacagcggcgt**cgc**gcgagcatacgg
Leu205Ala_F	Mutate *Mtb* UgpB residue Leu205Ala	gctaacgccgac**gcc**atctcgtggacg
Leu205Ala_R	Mutate *Mtb* UgpB residue Leu205Ala	cgtccacgagat**ggc**gtcggcgttagc
Trp208Ala_F	Mutate *Mtb* UgpB residue Trp208Ala	ccgacctcatctc**ggc**gacgtttcagggacc
Trp208Ala_R	Mutate *Mtb* UgpB residue Trp208Ala	ggtccctgaaacgtc**gcc**gagatgaggtcgg
Ser272Ala_F	Mutate *Mtb* UgpB residue Ser272Ala	gccgtggca**gcc**accggctcg
Ser272Ala_R	Mutate *Mtb* UgpB residue Gly306Ala	cgagccggt**ggc**tgccacggc
Gly306Ala_F	Mutate *Mtb* UgpB residue Ser272Ala	ccgacgggc**gct**gcggggctg
Gly306Ala_R	Mutate *Mtb* UgpB residue	cagccccgc**agc**gcccgtcgg

	Gly306Ala	
Tyr345Ala_F	Mutate *Mtb* UgpB residue Tyr345Ala	cagcaaaccggc**gct**ctgccggtgcgcaag
Tyr345Ala_R	Mutate *Mtb* UgpB residue Tyr345Ala	cttgcgcaccggcag**agc**gccggtttgctg
Arg385Ala_F	Mutate *Mtb* UgpB residue Arg385Ala	cacaagactacgca**gcg**gttttcctgcc
Arg385Ala_R	Mutate *Mtb* UgpB residue Arg385Ala	ggcaggaaaac**cgc**tgcgtagtcttgtg

Sequences in bold are the locations of the site-directed mutations.

Table 2.5: Empty vectors

Vector	Antibiotic resistance	Sequencing primer
pYUB1062	Hygromycin	T7_F and T7_R
pET160-DEST	Ampicillin	T7_F and T7_R
pET SUMO	Kanamycin	SUMO_F and T7_R
pYUB1062-GFP	Hygromycin	T7_F and pWaldo-R
pWaldo	Kanamycin	T7_F and pWaldo-R

Table 2.6: Constructs

Vector	Use	Reference
pYUB1062-UgpB	Expression of truncated UgpB (amino acid residues 1-34) in *M. smegmatis*	Constructed by Dr James Harrison, University of Warwick
pET160-DEST-UgpC	Expression of UgpC in *E. coli*	Glycerol stock
pET SUMO-UgpC	Expression of UgpC in *E. coli*	This study
pYUB1062-UgpC	Expression of UgpC in *M. smegmatis*	This study
pYUB1062-UgpC-GFP	Expression of UgpC GFP in *M. smegmatis*	This study
pWaldo-UgpC	Expression of UgpC GFP in *E. coli*	Glycerol stock
pWaldo-UgpAE	Expression of UgpAE GFP in *E. coli*	This study

Table 2.7: Bacterial strains

Strain	Genotype	Use
E. coli Top 10	F^- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ^- *rps*L(Str^R) *end*A1 *nup*G	General cloning strain Site-directed mutagenesis
E. coli DH5a	F^- φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_K^-, m_K^+) *pho*A *sup*E44 λ^- *thi*-1 *gyr*A96 *rel*A1	General cloning strain
E. coli BL21 (DE3)	F^- *omp*T *hsd*S_B (r_B^-, m_B^-) *gal dcm* (DE3)	Protein expression
E. coli BL21 (DE3) pLysS	F–, *omp*T, *hsd*S_B (r_B–, m_B–), *dcm*, *gal*, λ(DE3), pLysS, Cm^r.	Protein expression
E. coli Shuffle® T7	F´ *lac*, *pro*, *lacI*^{q} / *Δ(ara-leu)7697 araD139 fhuA2 lacZ::T7 gene1 Δ(phoA)PvuII phoR ahpC* galE (or U) galK λatt*::pNEB3-r1-*cDsbC*($Spec^R$, *lacI*^{q}) *ΔtrxB rpsL150*(Str^R) *Δgor Δ(malF)3*	Protein expression
E. coli Lemo21 (DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo*(Cam^R) *λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5* pLemo = pACYC184-*PrhaBAD-lysY*	Protein expression
M. smegmatis $mc^2$4517	$mc^2$155 F^- pYUB1232 acetamidase promoter T7 RNA polymerase (Wang et al., 2010)	Protein expression in *Mycobacterium smegmatis*

Table 2.8: Microbiological media

Media	Components
Luria-Bertani media (LB)	1 % (w/v) Tryptone (Sigma Aldrich) 0.5 % (w/v) yeast extract (Sigma Aldrich) 10 mM NaCl pH 7 in dH_2O
Luria-Bertani agar (LBA)	15 g/L Bacteriological agar (Sigma Aldrich) in LB media
Tryptic soy broth tween-80 (TSBT)	Tryptic soy broth powder 0.2 % (v/v) glycerol 0.05 % (v/v) tween-80 in dH_2O
Luria-Bertani glycerol tween-20 (LBGT)	LB broth powder 0.2 % (v/v) glycerol 0.05 % (v/v) tween-20 in dH_2O
Terrific broth (TB)	TB broth powder 1.0 % (v/v) glycerol
PASM-5052 media	2 mM $MgSO_4$ 0.2X trace metals 1X 5052 1X NPS 0.1 µM vitamin B12 100 µL 17aa 0.5 mg/mL methionine

Table 2.9: Antibiotics

	Bacterial species	
Antibiotic	***E. coli***	***M. smegmatis***
Ampicillin	100 µg/mL	100 µg/mL
Chloramphenicol	30 µg/mL	-
Kanamycin	50 µg/mL	25 µg/mL
Hygromycin	150 µg/mL	100 µg/mL

All antibiotics were dissolved in dH_2O apart from chloramphenicol that was dissolved in EtOH and filter sterilised using a 0.22 µm syringe (EMD Millipore) and stored at -20 °C.

Table 2.10: Thin layer chromatography (TLC) stains

Stain	Components	Detects
Sulphuric acid	10.5 mL sulphuric acid 95 % 89.5 mL ethanol 100 %	Carbohydrates
Ninhydrin	1.5 g ninhydrin (Sigma Aldrich) 100 mL n-butan-1-ol 3 mL acetic acid	Primary amines
Potassium permanganate	1.5 g $KMnO_4$ 10 g K_2CO_3 1.25 mL 10 % (w/v) NaOH 198.75 mL dH_2O	Oxidisable groups

Table 2.11: Commercially available crystallisation sparse matrix screens

Screen name	Supplier	Number of conditions
JCSG-*plus*™	Molecular Dimensions	96
Morpheus® HT-96	Molecular Dimensions	96
Structure Screen	Molecular Dimensions	96
MIDAS™	Molecular Dimensions	96
Additive Screen HT	Hampton Research	96

2.2 Vector maps

2.2.1 pYUB1062 vector

pYUB1062 (Wang et al., 2010) (Figure 2.1) is an *E. coli*-mycobacteria shuttle vector which is 4802 bp nucleotides long and contains a hygromycin resistance gene for selection purposes. The vector contains an *ori* for replication in *E. coli* and an f1 *ori* for replication in *Mycobacteria*. Expression of the insert is induced by the addition of acetamide in *Mycobacterium smegmatis* mc^{2}4517, a strain containing a copy of the T7 RNA polymerase integrated into its chromosome under the control of the acetamidase promoter. Acetamide induces T7 RNA polymerase production and expresses the insert under the control of the T7 promoter. The vector also has a hexa-histidine fusion tag region and in this study the gene cloned in this vector has a C-terminal tag for purification of the protein using immobilised metal affinity chromatography (IMAC).

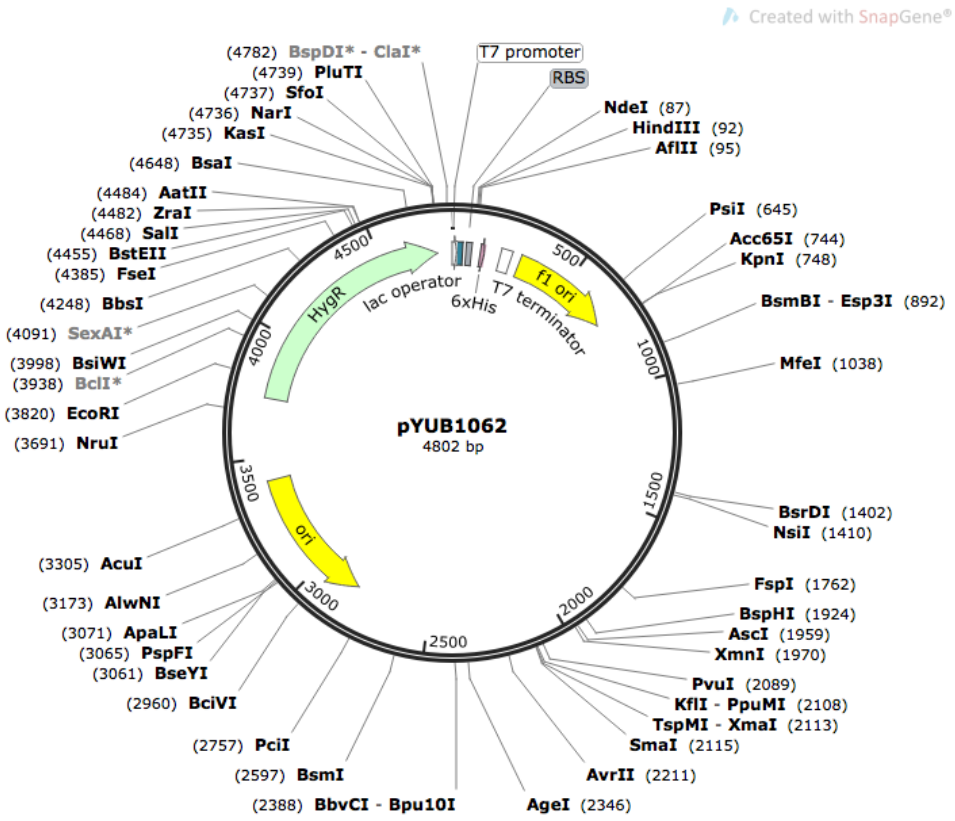


Figure 2.1: pYUB1062 vector map. Abbreviations: HygR hygromycin resistance gene, f1 ori and f1 ori origins of replication in *E. coli* and *Mycobacteria*.

2.2.2 pET160-DEST vector

pET160-DEST (Figure 2.2) is an *E. coli* gateway destination, expression vector which is 7437 bp nucleotides long and contains an ampicillin resistance gene for selection purposes. It has two recombination sites attR1 and attR2 downstream of the T7lac IPTG inducible promoter for recombination cloning. It has an N-terminal or C-terminal 6xHis tag for purification by immobilised metal affinity chromatography (IMAC).

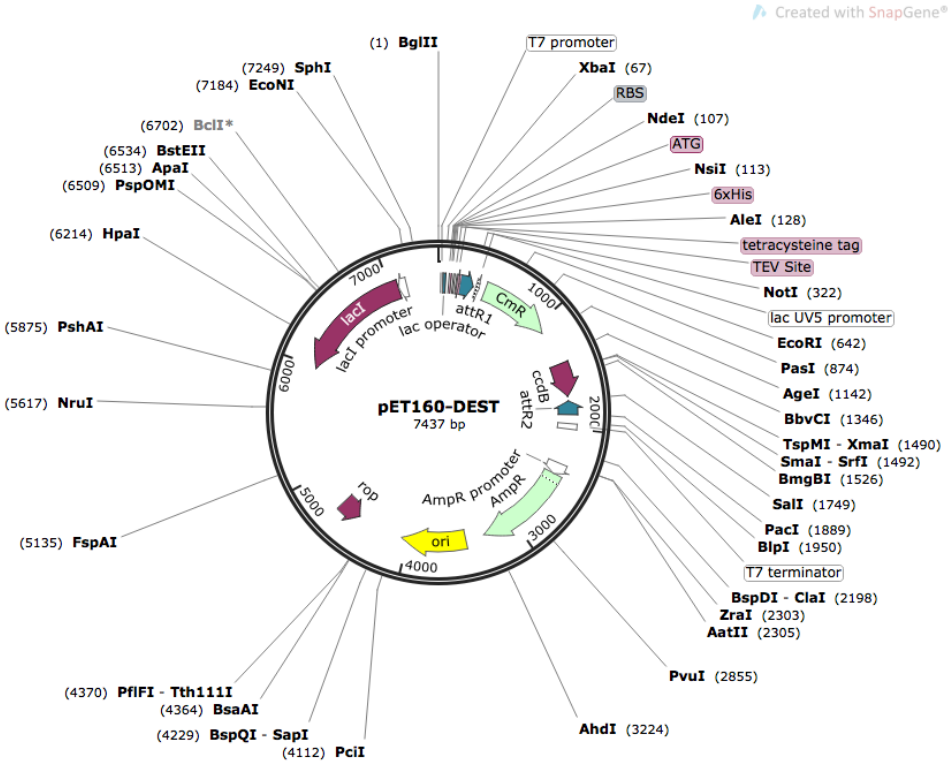


Figure 2.2: pET160-DEST vector map. Abbreviations: AmpR ampicillin resistance gene, ori origins of replication in *E. coli*.

2.2.3 pYUB1062-GFP vector

pYUB1062-GFP (Fullam et.al, 2018) (Figure 2.3) is a modified *E. coli*-mycobacterial shuttle vector based on pYUB1062 (Wang et al., 2010) which is 5542 bp nucleotides long. The modified vector has a C-terminal tobaccoetch virus (TEV) protease cleavage site and GFP fluorescent tag.

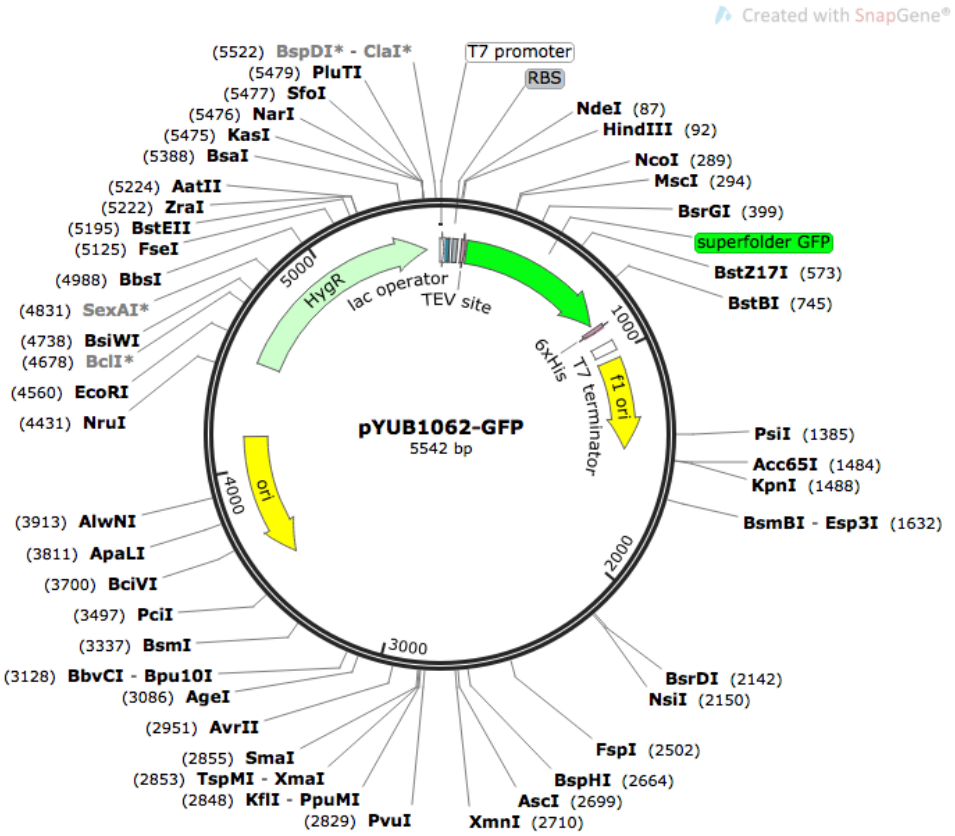


Figure 2.3: pYUB1062-GFP vector map (Fullam et.al, 2018)**.** Abbreviations: HygR hygromycin resistance gene, ori and f1 ori origins of replication in *E. coli* and *Mycobacteria*, GFP superfolder green fluorescent protein.

2.2.4 pET SUMO vector

pET SUMO (Invitrogen) (Figure 2.4) is an *E. coli* expression vector which is 5642 bp nucleotides long and contains a kanamycin resistance gene for selection purposes. The vector utilises the T7 promoter to control expression of the insert. Expression of the insert is induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) in BL21 (DE3) lysogen expression *E. coli* host strain. The vector has an N-terminal terminal hexa-histidine fusion tag (His.Tag®) and SUMO (*Saccharomyces cerevisiae* Smt3 protein) fusion solubility tag for purification purposes using immobilised metal affinity chromatography (IMAC).

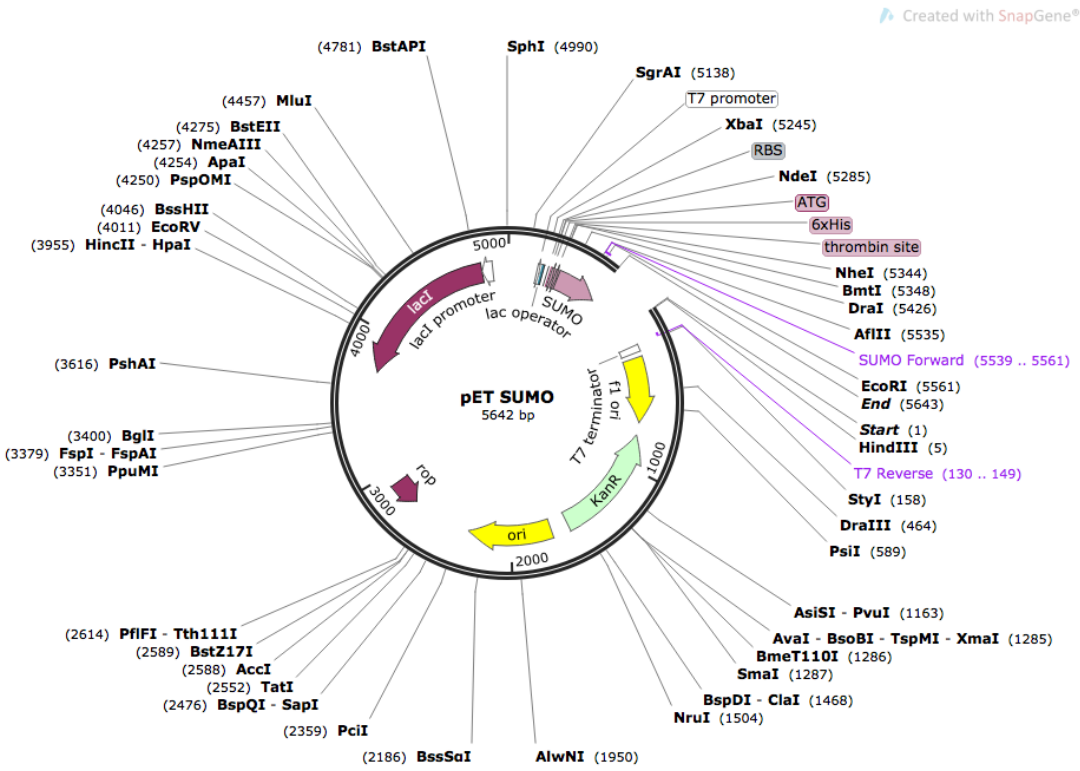


Figure 2.4: pET SUMO vector map (Invitrogen). Abbreviations: KanR kanamycin resistance gene, f1 ori origin of replication in *E. coli*, SUMO small ubiquitin-like modifier tag.

2.2.5 pWaldo vector

pWaldo (Drew et.al, 2001) (Figure 2.5) is an *E. coli* expression vector which is 6381 bp nucleotides long and contains a kanamycin resistance gene for selection purposes. Expression of the insert is induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) in BL21 (DE3) lysogen expression *E. coli* host strain. The vector is modified from pET28(a+) to contain a C-terminal tobaccoetch virus (TEV) protease cleavage site and GFP fluorescent tag. The GFP variant is selected to fold well in *E. coli* and has the red-shifted mutation S65T and the folding mutation F64L and a hexa-histidine fusion tag (His.Tag®) for purification purposes using immobilised metal affinity chromatography (IMAC).

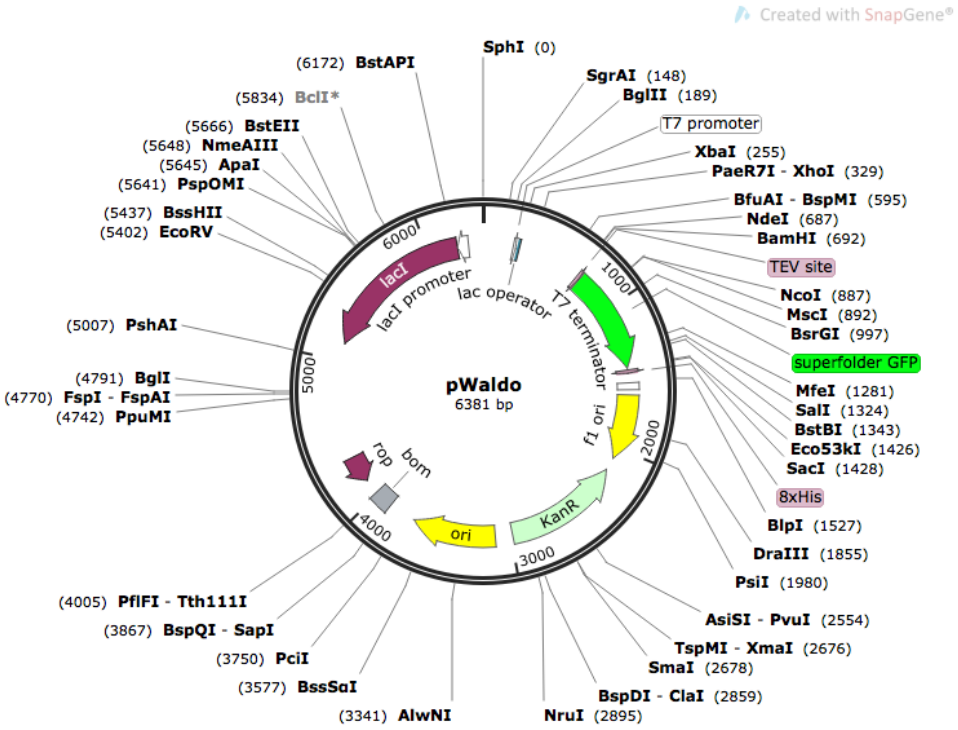


Figure 2.5: pWaldo vector map (Drew et.al, 2001). Abbreviations: KanR kanamycin resistance gene, f1 ori origin of replication in *E. coli*, GFP super folder green fluorescent protein.

2.3 Molecular biology

2.3.1 Agarose gel electrophoresis

Agarose gels were prepared using agarose 1 % (w/v) in TAE buffer (Table 2.1). 0.5 µM ethidium bromide was added to the dissolved agarose solution prior to casting of the gel. DNA loading dye (Thermo Scientific) was added to DNA samples and a 1 kbp DNA ladder (Thermo Scientific) was loaded onto the gel. The gels were electrophoresed at 120 V in TAE buffer and visualised using a UV transilluminator.

2.3.2 Determination of DNA concentration

DNA concentration was determined spectrophotometrically using a NanoDrop 2000 (Thermo Scientific) at 260 nm. The 260/280 nm ratio was also checked to assess DNA quality.

2.3.3 PCR amplification and cloning

2.3.3.1 PCR amplification and cloning of *ugpB* pYUB1062 from *M. tuberculosis*

Dr James Harrison carried out the following PCR and cloning. The UgpB gene was amplified from *M. tuberculosis* HR37 genomic DNA by PCR using primers UgpB_T_pYUB_F and UgpB_T_pYUB_R specific for *ugpB* (Table 2.2). The PCR reaction mixture contained Q5 DNA polymerase (NEB), Q5 DNA polymerase buffer, 0.2 mM dNTP's, 6 % (v/v) DMSO and 1 µM forward and reverse primers. The PCR cycle consisted of: denaturation (98 °C, 30 s) followed by 30 cycles of annealing (70 °C, 30 s), extension (72 °C, 90 s), followed by a final extension step (72 °C, 5 min).

The purified *ugpB* PCR product and pYUB1062 vector were digested with restriction endonucleases *NdeI* and *HindIII* (1 µL, 37 °C, 3 hours). The digested product was then ligated using T4 DNA ligase and ligase buffer (NEB) with a 3:1 ratio of vector: insert DNA (1 µL, 22 °C, 1 hour). The ligation product was transformed into *E. coli* Top10 and plated onto LB hygromycin (150 µg/mL) plates and left at 37 °C overnight until colonies appeared. Colonies were cultivated and the plasmid was isolated and sequenced with primers T7_F and T7_R (Table 2.3).

2.3.3.2 PCR amplification and cloning of *ugpC* pYUB1062 and *ugpC* pYUB1062-GFP from *M. tuberculosis*

The *ugpC* gene was amplified from *M. tuberculosis* HR37 genomic DNA by PCR using the following primers specific for UgpC: (UgpC pYUB1062_F and UgpC pYUB1062_R) or (UgpC pYUB1062_F and UgpC pYUB1062 GFP_R) (Table 2.2). The PCR reaction

mixture contained Q5 DNA polymerase (NEB), Q5 DNA polymerase buffer, 0.2 mM dNTP's, 6 % (v/v) DMSO and 0.5 µM forward and reverse primers. The PCR cycle consisted of: denaturation (98 °C, 30 s) followed by 30 cycles of annealing gradient (64-80 °C, 30 s), extension (72 °C, 90 s), followed by a final extension step (72 °C, 5 min).

The purified *ugpC* PCR product, pYUB1062 and pYUB1062-GFP vectors were digested with restriction endonucleases *NdeI* and *HindIII* (1 µL, 37 °C, 3 hours). The digested product was then ligated using T4 DNA ligase and ligase buffer (NEB) with a 3:1 ratio of vector: insert DNA (1 µL, 22 °C, 1 hour). The ligation product was transformed into *E. coli* Top10 and plated onto LB hygromycin plates and left at 37 °C overnight until colonies appeared. The presence of the insert was determined by colony PCR using primers (UgpC pYUB1062_F and UgpC pYUB1062_R) or (UgpC pYUB1062_F and UgpC pYUB1062 GFP_R). The plasmid was isolated from colonies containing the insert and sequenced with the primers T7_F and T7-R for UgpC pYUB1062 and T7_F and pWaldo-R for UgpC pYUB1062-GFP (Table 2.3).

2.3.3.3 PCR amplification and cloning of *ugpC* pETSUMO from *M. tuberculosis*

The *ugpC* gene was amplified from *M. tuberculosis* HR37 genomic DNA by PCR using the following primers specific for UgpC: UgpC pET SUMO_F and UgpC pET SUMO_R (Table 2.2). The PCR reaction mixture contained Q5 DNA polymerase (NEB), Q5 DNA polymerase buffer, 0.2 mM dNTP's, 6 % (v/v) DMSO and 0.5 µM forward and reverse primers. The PCR cycle consisted of: denaturation (98 °C, 30 s) followed by 30 cycles of annealing gradient (68-80 °C, 30 s), extension (72 °C, 90 s), followed by a final extension step (72 °C, 5 min).

The purified *ugpC* PCR product and pETSUMO vectors were digested with restriction endonucleases *BamHI* and *HindIII* (1 µL, 37 °C, 3 hours). The digested product was then ligated using T4 DNA ligase and ligase buffer (NEB) with a 3:1 ratio of vector: insert DNA (1 µL, 22 °C, 1 hour). The ligation product was transformed into *E. coli* Top10 and plated onto LB kanamycin plates and left at 37 °C overnight until colonies appeared. The presence of the insert was determined by colony PCR using primers UgpC pET SUMO_F and UgpC pET SUMO_R (Table 2.2) The plasmid was isolated from colonies containing the insert and sequenced with the primers SUMO_F and T7_R (Table 2.3).

2.3.3.4 PCR amplification and cloning of *ugpAE* from *M. tuberculosis*

The UgpAE gene was amplified from *M. tuberculosis* HR37 genomic DNA by PCR using primers specific for UgpAE: UgpAE pWaldo_F and UgpAE pWaldo_R (Table 2.2). The PCR reaction mixture contained Q5 DNA polymerase (NEB), Q5 DNA polymerase buffer, 0.2 mM dNTP's, 6 % (v/v) DMSO and 0.5 µM forward and reverse primers. The PCR cycle consisted of: denaturation (98 °C, 30 s) followed by 30 cycles of annealing gradient (72-82 °C, 30 s), extension (72 °C, 90 s), followed by a final extension step (72 °C, 5 min).

The purified *ugpAE* PCR product, and pWaldo vector were digested with restriction endonucleases *NdeI* and *BamHI* (1 µL, 37 °C, 3 hours). The digested product was then ligated using T4 DNA ligase and ligase buffer (NEB) with a 3:1 ratio of vector: insert DNA (1 µL, 22 °C, 1 hour). The ligation product was transformed into *E. coli* Top10 and plated onto LB kanamycin plates and left at 37 °C overnight until colonies appeared. The presence of the insert was determined by colony PCR using primers UgpAE pWaldo_F and UgpAE pWaldo_R (Table 2.2). The plasmid was isolated from colonies containing the insert and sequenced with primers T7_F and pWaldo-R (Table 2.3).

2.3.4 DNA purification

Plasmid DNA samples were purified from overnight bacterial cultures of *E. coli* using a QIAprep spin mini prep kit (Qiagen). PCR products and enzymes reactions were purified using a QIAquick PCR purification kit (Qiagen). DNA purified from agarose gels was purified using a QIAquick gel extraction kit (Qiagen). All kits were used according to the manufacturers instructions.

2.3.5 Restriction endonucleases

Purified DNA samples were digested with restriction endonucleases (NEB) typically using 10U per µg of DNA at 37 °C. Endonuclease reactions were purified using a QIAquick PCR purification kit (Qiagen). The digested vectors were dephosphorylated with calf-intestinal alkaline phosphatase (NEB) using 1U per pmol of DNA at 37 °C for 30 min to dephosphorylate the 3' ends.

2.3.6 Site-directed mutagenesis of *Mtb ugpB*-pYUB1062

The pYUB1062 plasmid containing the *ugpB* insert was isolated and purified from *E. coli* Top 10 cells. Targeted single-site substitutions were introduced into the *ugpB* gene using forward and reverse primers specific for the substitutions (Table 2.4). Site-directed mutagenesis was carried out by PCR using Phusion high fidelity DNA

polymerase (NEB). The PCR cycle consisted of: denaturation (98 °C, 30 s) followed by 20 cycles of annealing at a temperature 5 °C below the primer melting temperature (30 s), extension (72 °C, 90 s), followed by a final extension step (72 °C, 5 min). 1 µL of the restriction endonuclease *DpnI* (NEB) was added and the PCR reaction was incubated at 37 °C for 3 h and transformed into *E. coli*.

2.3.7 Transformation of *E. coli* via heat shock

The ligation mixtures or *DpnI* digested mutated plasmid (1-2 µL) was added to 50 µL of thawed *E. coli* Top 10 cells at 4 °C in pre-chilled sterile 14 mL ventilated culture tubes, gently mixed and incubated on ice for 30 min. The transformation mix was heated to 42 °C for 45 s and then incubated on ice for 5 min. LB media (200 µL) was added and the bacterial culture grown at 37 °C for 1 h with shaking at 180 rpm and then spread onto LB agar plates supplemented with the appropriate antibiotics. The plates were incubated at 37 °C for 16 h. The colonies were screened for the presence of the correct plasmid containing insert by PCR and the plasmid containing the insert was extracted and confirmed by sequencing.

2.3.8 Sequencing plasmids

Individual colonies containing plasmid were picked, and incubated in 4 mL of LB in ventilated culture tubes supplemented with the appropriate antibiotics at 37 °C overnight with shaking at 180 rpm overnight. The plasmid was extracted using a mini-prep kit (Qiagen). The DNA was sent for sequencing (GATC Biotech) using the specific sequencing primers (Table 2.3).

2.3.9 Storage of transformed *E. coli* cells

E. coli cells containing the confirmed constructs were stored at -80 °C as glycerol stocks. A single colony containing the correct sequenced insert was picked from an LB agar plate and used to inoculate 4 mL of LB media containing the appropriate antibiotics. The cultures were grown at 37 °C with shaking at 180 rpm until the culture reached mid-log phase (OD_{600nm} = 0.6 – 0.8). Sterile glycerol was added to the culture to give a final concentration of 10 % (v/v), divided into 1 mL aliquots in 2 mL cryovials (Nalgene) and stored at -80 °C.

2.3.10 Preparation of competent *M. smegmatis* $mc^2$4517

A glycerol stock of *M. smegmatis* $mc^2$4517 was streaked into LB plates and incubated at 37 °C for 2 -3 days until colonies appeared. A single colony was picked and used to inoculate 4 mL of TSBT media in 14 mL ventilated culture tubes and incubated at 37 °C for 2-3 days with shaking at 180 rpm. This culture was used to inoculate 50 mL of

TSBT in a 1:100 dilution until OD_{600nm} = 0.8-1.0 was reached. The cells were incubated at 4 °C for 30 min and centrifuged at 3,000 g for 15 min. The cells were resuspended in 10 mL of ice-cold 10 % glycerol at 4 °C and centrifuged at 3,000 g for 15 min. The supernatant was discarded and cells were resuspended in 5 mL 10 % glycerol at 4 °C, aliquoted into 0.2 mL and stored at -80 °C.

2.3.11 Electroporation of *M. smegmatis* mc^{2}4517

0.5-5 µg of DNA was added to 0.2 mL electrocompetent *M. smegmatis* MC^{2}4517 cell aliquots and transferred to 0.2 mL electroporation cuvettes (Biorad). After incubation on ice for 10 min the cuvette was placed in the electroporator and pulsed at 2.5 kV. The cells were diluted with 2 mL LBGT media and incubated at 37 °C for 3 h without antibiotics. Cells were then plated onto LB supplemented with the appropriate antibiotics and incubated for 72 h at 37 °C until colonies appeared.

2.4 Protein expression and purification

2.4.1 Expression of *M. tuberculosis* UgpB-pYUB1062 in *M. smegmatis* and genetic variants

A single colony of *M. smegmatis* $mc^2$4517 UgpB-pYUB1062 was used to inoculate 100 mL of LBGT media containing the appropriate antibiotics (25 µg/mL of kanamycin and 100 µg/mL hygromycin). The starter culture was incubated for 72 h at 37 °C with shaking at 180 rpm. 14 mL of starter culture was used to inoculate 6 L of LBGT media to give a starting optical density (OD_{600nm}) of 0.1. The culture was incubated at 37 °C with shaking at 180 rpm until the OD_{600nm} reached 0.4. Gene expression was induced by the addition of 0.2 % (w/v) acetamide and the culture was incubated at 37 °C for 20 h. Cells were centrifuged at (10,000 g, 40 min, 4 °C) and washed by re-suspending in phosphate buffered saline (PBS) and pelleted again (3,000 g, 40 min, 4 °C) prior to freezing at -80 °C.

2.4.2 Expression of *M. tuberculosis* UgpC-pET SUMO in *E. coli*

A single colony of UgpC-pET SUMO *E. coli* BL21 (DE3) was picked and inoculated in a 50 mL starter culture of LB kanamycin (50 µg/mL) at 37 °C with shaking overnight. The starter culture was used to inoculate a 1:100 dilution of 1 L terrific broth containing kanamycin (50 µg/mL) and incubated at 27 °C with shaking at 180 rpm until OD_{600nm} of 0.6 was reached. 1 mM of IPTG was added and the temperature was shifted to 16 °C overnight. Cells were centrifuged at (10,000 g, 40 min, 4 °C) and washed by re-suspending in phosphate buffered saline (PBS) and pelleted again (3,000 g, 40 min, 4 °C) prior to freezing at -80 °C.

2.4.3 Expression of *M. tuberculosis* UgpC-pYUB1062/pYUB1062-GFP in *M. smegmatis*

A single colony of *M. smegmatis* $mc^2$4517 containing either UgpC-pYUB1062 or UgpC-pYUB1062-GFP was used to inoculate 100 mL of LBGT media containing the appropriate antibiotics (25 µg/mL of kanamycin and 100 µg/mL hygromycin). The starter culture was incubated for 72 h at 37 °C with shaking at 180 rpm. 14 mL of starter culture was used to inoculate 1 L of LBGT media to give a starting optical density (OD_{600nm}) of 0.1. The culture was incubated at 37 °C with shaking at 180 rpm until the OD_{600nm} reached 0.4. Gene expression was induced by the addition of 0.2 % (w/v) acetamide and the culture was incubated at 37 °C for 20 h. After this time, 1 mL of the UgpC-pYUB1062-GFP cells was centrifuged (10,000 g, 10 min) and resuspended in 100 µL PBS and GFP fluorescence was measured at λ_{ex} 485 nm λ_{em} 535 nm in a plate reader (Tecan Infinite F200, gain 35). The remaining 1 L culture was

centrifuged at (10,000 g, 40 min, 4 °C) and washed by re-suspending in phosphate buffered saline (PBS) and pelleted again (3,000 g, 40 min, 4 °C) prior to freezing at -80 °C.

2.4.4 Expression of *M. tuberculosis* UgpC and UgpAE-pWaldo in *E. coli*

A single colony of *E. coli* Lemo21 (DE3) containing either UgpC-pWaldo or UgpAE-pWaldo was used to inoculate 5 mL of PASM-5052 media (Table 2.1) containing the appropriate antibiotics (30 µg/mL of chloramphenicol and 50 µg/mL kanamycin). The starter culture was incubated overnight at 37 °C with shaking at 200 rpm. 50 µL of the starter culture (1: 100 ratio dilution) was used to inoculate 5 mL of PASM-5052 media and the culture was incubated at 37 °C with shaking at 200 rpm until the OD_{600nm} reached 0.6. Gene expression was induced by the addition of 0.4 mM IPTG and 6 concentrations of a L-rhamnose gradient (0 – 1 mM final concentration) and the cultures were incubated at 25 °C with shaking at 200 rpm for 15 h. After induction the OD_{600nm} was measured and 1 mL of the cells were centrifuged (10,000 g, 10 min) and resuspended in 100 µL PBS and GFP fluorescence was measured at λ_{ex} 485 nm λ_{em} 535 nm in a plate reader (Tecan Infinite F200, gain 35). Cells were centrifuged at (3,000 g, 40 min, 4 °C) and washed by re-suspending in phosphate buffered saline (PBS) and pelleted again (3,000 g, 40 min, 4 °C) prior to freezing at -80 °C.

2.4.5 Cell lysis

Cell pellets were thawed at 4 °C and re-suspended in 20 mL of the appropriate lysis buffer (Table 2.1) supplemented with one EDTA-free protease inhibitor cocktail tablet (Pierce) and 1 µg/mL DNAse (Roche). The cell suspensions were ultrasonicated at 4 °C (Misonix) at 70 % power for 20 cycles of 30 s ON and 30 s OFF and centrifuged (12,000 g, 1h, 4 °C) to remove insoluble cell debris. The supernatant was filtered through a 0.45 µm filter (EMD Millipore) prior to immobilised metal affinity chromatography (IMAC).

2.4.6 Purification of *M.* tuberculosis UgpB-pYUB1062

Purification of recombinant 6x his tagged proteins was carried out using cobalt immobilised metal affinity chromatography. A gravity flow column (PD20 BioRad) was prepared by the addition of a 2 mL bed volume of HisPur™ cobalt chromatography resin (Thermo Scientific). The resin was washed twice with 10 mL of dH_2O and equilibrated in 20 mL of UgpB buffer 1 (Table 2.1). The filtered cell lysate was added to the column and the flow through was collected and passed over the column again. The bound lysate was washed with UgpB wash buffer (Table 2.1) followed by UgpB buffer 1

containing increasing concentration of imidazole from 5 mM, 10 mM, 20 mM, 50 mM, 100 mM, 250 mM, 500 mM and 1 M to elute the bound proteins. Samples of collected fractions were subjected to SDS-PAGE and samples that contained UgpB or the genetic mutants of UgpB were combined and dialysed (12-14 kDa MWCO Medicell Membranes) overnight at 4 °C in UgpB dialysis buffer 1 (Table 2.1) which was replaced the following day twice over 4 h.

After dialysis the protein was filtered (0.22 µm EMD Millipore). A 5 mL HiTrap QHP ion exchange chromatography column (GE Healthcare) was equilibrated in UgpB dialysis buffer 1 (Table 2.1). The protein was passed over the column at approximately 1 mL/min and the flow-through was collected and re-passed over the column. A step-wise salt gradient was performed from 150 mM – 1 M NaCl in UgpB dialysis buffer 1, 5 mL (Table 2.1) to elute the bound proteins. Fractions were collected and subjected to SDS-PAGE.

Fractions containing the UgpB protein at the expected size were combined and concentrated by ultrafiltration (Vivaspin 10 kDa MWCO) to 2 mL prior to loading onto the AKTA pure system connected to a HiLoad 16/600 Superdex 75 pg gel filtration chromatography column (GE Healthcare). The HiLoad 16/600 Superdex 75 pg gel filtration chromatography column was first equilibrated in UgpB dialysis buffer 2 (Table 1) and the proteins were separated at 0.5 mL/min and the absorbance was monitored at 280 nm. The chromatogram sample peak fractions were subjected to SDS-PAGE and fractions containing pure UgpB protein at the expected size were combined, and concentrated to ~ 2-10 mg/mL by ultrafiltration (Vivaspin 10 kDa MWCO) prior to flash freezing and storage at -80 °C.

2.4.7 Purification of *M. tuberculosis* UgpC-pET SUMO

The cell pellet was thawed and re-suspended in 20 mL of UgpC lysis buffer (Table 2.1) supplemented with 0.1 % Triton-X 100 and one EDTA-free protease inhibitor cocktail tablet (Pierce). The cell suspension was ultrasonicated (Misonix) at 70 % power for 10 cycles of 30 s ON and 30 s OFF and centrifuged (12,000 g, 1h, 4 °C) to remove insoluble cell debris. The supernatant was filtered through a 0.45 µm filter (EMD Millipore) prior to nickel immobilised metal affinity chromatography (IMAC). UgpC was eluted from the Ni^{2+} column in UgpC lysis buffer (Table 2.1) with a step-wise imidazole gradient. Fractions containing the protein were dialysed to remove imidazole and 150 µL SUMO protease (0.3 mg/mL) was added and incubated at 30 °C for 30 min. The digested protein was loaded onto another Ni^{2+} column and eluted from the column in

UgpC lysis buffer with a step-wise imidazole gradient. The flow through and wash were combined, concentrated to 2 mL (Vivaspin 10 kDa MWCO) prior to loading onto the AKTA pure system connected to a HiLoad 16/600 Superdex 200 pg gel filtration chromatography column (GE Healthcare). The HiLoad 16/600 Superdex 200 pg gel filtration chromatography column column was first equilibrated in UgpC lysis buffer (Table 2.1) and the proteins were separated at 0.5 mL/min and the absorbance was monitored at 280 nm. The chromatogram sample peak fractions were subjected to SDS-PAGE and fractions containing pure UgpB protein at the expected size were combined, and concentrated to ~ 2-10 mg/mL by ultrafiltration (Vivaspin 10 kDa MWCO) prior to flash freezing and storage at -80 °C.

2.4.8 Attempts to purify *M. tuberculosis* UgpC-pYUB1062/pYUB1062-GFP

Purification of recombinant 6x his tagged proteins was carried out using nickel immobilised metal affinity chromatography. A gravity flow column (PD20 BioRad) was prepared by the addition of a 2 mL bed volume of HisPur™ nickel-NTA chromatography resin (Thermo Scientific). The resin was washed twice with 10 mL of dH_2O and equilibrated in 20 mL of UgpC buffer (Table 2.1). The filtered cell lysate was added to the column and the flow through was collected and passed over the column again. The bound lysate was washed with UgpC buffer (Table 2.1) followed by UgpC buffer containing increasing concentration of imidazole from 5 mM, 10 mM, 20 mM, 50 mM, 100 mM, 250 mM, 500 mM and 1 M to elute the bound proteins. Samples of collected fractions were subjected to SDS-PAGE and in-gel GFP fluorescence was measured.

2.4.9 Attempts to purify *M. tuberculosis* UgpC/UgpAE-pWaldo

Purification of recombinant 6x his tagged proteins was carried out using nickel immobilised metal affinity chromatography. A gravity flow column (PD20 BioRad) was prepared by the addition of a 2 mL bed volume of HisPur™ nickel-NTA chromatography resin (Thermo Scientific). The resin was washed twice with 10 mL of dH_2O and equilibrated in 20 mL of pWaldo purification buffer (Table 2.1). The filtered cell lysate was added to the column and the flow through was collected and passed over the column again. The bound lysate was washed with pWaldo purification buffer (Table 2.1) followed by UgpC buffer containing increasing concentration of imidazole from 5 mM, 10 mM, 20 mM, 50 mM, 100 mM, 250 mM, 500 mM and 1 M to elute the bound proteins. Samples of collected fractions were subjected to SDS-PAGE and in-gel GFP fluorescence was measured.

2.4.10 Concentration of protein samples

Proteins were concentrated by ultrafiltration (Vivaspin 10 kDa MWCO) prior to flash freezing and storage at -80 °C.

2.5 Protein characterisation

2.5.1 SDS-PAGE

Samples were prepared by adding 2X SDS-loading dye to protein samples (Table 2.1) and heating to 95 °C for 5 min. Electrophoresis was performed at 50 mA for 33 min in 1X gel running buffer (Table 2.1) using precast polyacrylamide gels (BioRad AnykD). PageRuler prestained protein ladder (NEB) was used as a molecular weight reference. Gels were stained in InstantBlue™ (Expedeon) and de-stained overnight in dH_2O.

2.5.2 Determination of protein concentration

Protein concentration was determined spectrophotometrically using a NanoDrop 2000 (Thermo Scientific) at 280 nm.

2.5.3 GFP fluorescence

Prior to harvesting the cells, 1 mL of the culture was removed, centrifuged (2,600 g, 10 min, 4 °C), the pellet resuspended in 100 µL PBS and the fluorescence monitored at λ_{ex} 485 nm λ_{em} 535 nm (Tecan Infinite F200, gain 35).

2.5.4 In-gel GFP fluorescence

Samples for SDS-PAGE were mixed in a 1:1 ratio with SDS-loading buffer and 15 µL samples were loaded directly onto the gel without heating. Following SDS-PAGE (2.5.1) however without staining with InstantBlue™ the in-gel fluorescence was imaged immediately using λ_{ex} 473 nm [LBP], 450 V PMT, 50-µm-pixel size (Typhoon™ FLA 9500). The gels were then stained with InstantBlue™.

2.5.5 In-gel proteomic analysis

Gel slices containing the protein of interest were cut out using a clean scalpel, diced into 2-4 mm cubes and transferred to a 1.5 mL tube. The gel dices were de-stained by addition of 500 µL of gel wash buffer (Table 2.1) and heated at 55 °C for 20 min with shaking 650 rpm. The wash step was repeated 3 times. The gel dices were dehydrated by addition of 150 µL of 100 % ethanol and incubated at room temperature for 5 min with shaking 650 rpm. 150 µL of reducing buffer (Table 2.1) was then added to reduce the disulphide bonds and the gel dices were heated at 55 °C for 30 min with shaking 650 rpm. The liquid was removed and 150 µL of alkylation buffer (Table 2.1) was added, the tube was covered with foil and left at room temperature for 20 min to alkylate the free cysteine residues. The gel dices were washed 2 times with 500 µL gel wash buffer (Table 2.1) and incubated at room temperature for 20 min with shaking 650 rpm. The gel dices were dehydrated by the addition of 150 µL of 100 % ethanol

and incubated at room temperature for 5 min with shaking 650 rpm, rehydrated by the addition of 60 μL of trypsin buffer (Table 2.1), covered with aluminium foil and digested overnight at 37 °C. Peptides were extracted 2 times by the addition of 120 μL of 25 % (v/v) acetonitrile and 5 % (v/v) formic acid and sonicated for 10 min. The peptide extract was centrifugal filtered (Costar 0.22 μm), transferred to a clean tube and concentrated (SpeedVac) to a volume of 20 μL. Peptides were re-suspended to 50 μL in 2.5 % (v/v) acetonitrile and 0.05 % (v/v) trifluoroacetic acid.

A 20 μL aliquot was analysed by nanoLC-ESI-MS/MS using the Ultimate 3000/Orbitrap Fusion instrument (Thermo Scientific) specifying a 60-min LC separation on a 50 cm column. Data were used to interrogate the *Mycobacterium tuberculosis*, *Mycobacterium smegmatis* and *E.coli* proteome databases (UniProt) and the common contaminant database (MaxQuant) to identify and quantify any proteins present. The results were presented and analysed (Scaffold).

2.5.6 Circular dichroism analysis

Purified *Mtb* UgpB proteins were dialysed (12-14 kDa MWCO Medicell Membranes) into low salt phosphate buffer (Table 2.1), diluted to 0.25 mg/mL and transferred to a 1 mm path length quartz cuvette. The cuvette was cleaned between analyses with 2 M HNO_3, dH_2O, 1 % (v/v) Helmanex III (Sigma Aldrich) and dH_2O followed by drying with N_2. Circular dichroism spectra were acquired on a Jasco J-810 DC spectrometer from 198-260 nm in triplicate and averaged after subtraction of the buffer background.

2.5.7 Western blot

SDS-PAGE gels were run (2.5.1) and then transferred to a nitrocellulose membrane (Amersham Hybond ECL) (1 h, 20 V, 300 mA) in transfer buffer (Table 2.1). The membrane was blocked in 5 % milk (w/v) TBS-tween for 60 min at room temperature with shaking**.** Penta-His antibody, BSA-free (Sigma) was prepared (1 in 1000 dilution) in 25 mL TBS-tween containing 0.1 % milk (w/v). The 5 % milk TBS-tween was replaced with Penta-His antibody solution and incubated for 30 min at room temperature with shaking. The membrane was washed with 50 mL TBS-tween for 5 min three times. 1 μL of Anti-mouse IgG-Alkaline Phosphatase secondary antibody (Sigma) was added to 25 mL TBS-tween containing 0.1 % milk (w/v) and incubated with the membrane for 30 min at room temperature with shaking. The membrane was washed with 50 mL TBS-tween for 5 min three times and a final 50 mL of TBS. One SigmaFast BCIP/NBT tablet was dissolved in 10 mL of dH_2O and added to the membrane and allowed to stain for 5 min until bands appeared.

2.5.8 Thermal shift assay

The transition unfolding temperature T_m of the UgpB protein (22 µM) was measured in the presence or the absence of ligands. The screen used a single ligand concentration of 100 mM. Measurements were performed in a total volume of 20 µL using a Rotor-Gene Q Detection System (Qiagen), setting the excitation wavelength to 470 nm and detecting emission at 557 nm of the SYPRO Orange protein gel stain, 15 X final concentration (Invitrogen). The following cycle was used: melt ramp from 30 °C to 95 °C, increasing temperature in 1 °C steps and time intervals of 5 s. Fluorescence intensity was plotted as a function of the temperature and the T_m was determined using the Rotor-Gene Q software and Analysis Melt functionality. All experiments were performed in triplicate.

2.5.9 Phospho and sphingo lipid strips

Phosphoinositide strip, phosphoinostide array and sphingo strip membranes (Echelon Biosciences) were probed according to the manufacturers guidance. The membranes were transferred to 50 mL falcon tubes and blocked at 4 °C overnight with rolling in 5 mL of TBST (Table 2.1) supplemented with 3 % fatty acid free bovine serum albumin (Sigma Aldrich). The TBST blocking buffer was replaced and *Mtb* UgpB (5 µg/mL) was added and incubated for 2 h at room temperature with rolling. The membranes were washed 6 times with 5 mL of TBST every 5 min for 30 min after all incubation steps. Before and during the antibody detection steps the membranes were blocked in TBST (Table 2.1) supplemented with 1 % (w/v) fatty acid free bovine serum albumin (Sigma Aldrich). Bound UgpB was detected by incubating with Penta-His primary antibody (Qiagen) diluted 1:1000 for 1 h at room temperature followed by washes and incubating with anti-mouse IgG-HRP secondary antibody (Santa Cruz Biotech) diluted 1:2000 for 1 h at room temperature. Chemiluminescence using enhanced chemiluminescence reagent (ECL Pierce™) was used to detect the bound antibody by covering the membrane with 1 mL of ECL and incubating 5 min and imaged using an ImageQuant LAS 500 imager set to record images incrementally every 30 s.

The membranes were quantified by calculating the integrated density of spots in ImageJ (Fiji). The 8-min incremental exposure time images were used in all quantifications. A 52 by 52-pixel circle was specified to measure the integrated density of the lipid spots. The solvent blank was subtracted from all density values and the densities were presented relative to the highest binder set arbitrarily to 100.

2.5.10 Microarray array

All analyses were performed at ambient temperature and microarray slides were initially blocked with 10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$ containing 1 % BSA and 0.02 % casein for 1 hour. For the non-precomplexed experiment, *Mtb* UgpB protein was diluted in the blocking solution to (100 μg/mL) and overlaid for 1.5 hours, followed by detection with primary monoclonal anti-polyHistidine antibody and secondary biotinylated anti-mouse IgG at (10 μg/mL) each incubated for 1 hour. For the pre-complex experiment, protein-antibodies complexes were prepared by pre-incubating primary antibody and secondary antibody for 15 minutes, followed by addition of *Mtb* UgpB protein and incubation for a further 15 minutes, the ratio of His-tagged protein: primary antibody: secondary antibody was 1:2:2 (w/w/w). After diluting in the blocking solution, the complexes were overlaid at 25 μg/mL (concentration of *Mtb* UgpB) for 1.5 hour. For both the non-precomplexed and precomplexed protein, binding was detected after 30 minutes overlay with streptavidin-Alexa Fluor 647 (1 μg/mL).

2.5.11 Bio-Layer interferometry

Biolayer interferometry was carried out on a ForteBio Octet Red 96 instrument (Forte Bio). Assays were performed in black 96 well half-area plates (Greiner bio-one). Assays were carried out at room temperature using Bio-Layer buffers (Table 2.1) keeping the same particular buffer throughout for individual experiments. UgpB protein was first dialysed into the appropriate Bio-Layer buffer and diluted 5-20 μg/mL. EDC-NHS biosensor tips (Forte Bio) were hydrated in dH_2O for at least 10 min prior to use and the biosensor tip agitation speed was set to 1,000 rpm. The tips were functionalised by loading UgpB (5-20 μg/mL) and the experimental setup was the following: 1. dH_2O (equilibration 0-60 s), 2. EDC (20 mM) NHS (10 mM) (activation 60-360 s), 3. UgpB (loading 360-960 s), 4. Ethanolamine (1M) (quenching 960-1260 s), 5. buffer (buffer baseline 1260-1380 s), 6. glycerol-3-phoshocholine (association 1380-1675 s) and 7. buffer (dissociation 1675-1980 s).

2.5.12 Microscale thermophoresis (MST)

Mtb UgpB protein was labelled using Monolith His-tag labelling kit RED-Tris-NTA (NanoTemper Technologies) according to the manufactures guidance. PBST (Table 2.1) and a constant concentration of UgpB (50 nM) were used throughout. The ligands to test were prepared in PBST in the concentration range 0-0.5 M. Samples were loaded into the MonoLith NT.115 standard treated capillaries and incubated for 10 min before analysis using the Monolith NT.115 instrument (NanoTemper Technologies) at 21 °C using medium laser power and 40 % LED power. The binding affinities were

calculated by fitting a K_d model using the in-built MST analysis software and curves were plotted in GraphPad (Prism Version 7.0). All experiments were carried out in triplicate.

2.5.13 Malachite green phosphate assay

A malachite green phosphate assay kit (Sigma Aldrich) was used according to the manufacturers instructions. UgpC protein was washed extensively with assay buffer (Table 2.1) using a Vivaspin 10 kDa MWCO centrifugal concentrator and a final concentration of 0.6 mg/mL of UgpC protein was used in the assay. ATP (0.1 mM) and $MgCl_2$ (1 mM) were added in each 80 μL assay. Reactions were incubated at 37 °C, 300 rpm and quenched at different time points by the addition of 20 μL of freshly prepared working reagent (Sigma Aldrich), incubated at room temperature for 30 min prior to measurement of absorbance at 595 nm in a plate reader.

2.6 Protein structure determination

2.6.1 Methylation of *Mtb* UgpB

Purified *Mtb* UgpB was diluted to 1 mg/mL in UgpB dialysis buffer 2 (Table 2.1) and reductively methylated following a previous study (Walter et al., 2006). Briefly, 1 M dimethylamine borane complex (DMAB) and 1 M formaldehyde were dissolved in UgpB dialysis buffer 2. DMAB and formaldehyde were added to 280 μL of *Mtb* UgpB (1 mg/mL) to a final concentration of 20 mM DMAB and 40 mM formaldehyde. The mixture was incubated (100 rpm, 4 ^{o}C, 2 h) and after 2 h incubation, 6 μL of DMAB (1 M) and 12 μL of formaldehyde (1 M) was added and incubated (100 rpm, 4 ^{o}C for 2 h). This step was repeated twice. Tris-HCl pH 7.0 was added to a final concentration of 100 mM to remove any excess unreacted DMAB. The methylated protein was dialysed in 2 L of UgpB dialysis buffer 2 (4^{o} C, 16 h) and dialysis was repeated the following day (4 ^{o}C, 4 h). Methylated *Mtb* UgpB was then concentrated by ultrafiltration (Vivaspin 10 kDa MWCO) to 7 mg/mL.

2.6.2 Crystallisation and structure determination GPC bound *Mtb* UgpB

For co-crystallisation experiments methylated *Mtb* UgpB was incubated with 10 mM glycerophosphocholine (GPC) (Carbosynth) and incubated at 4 °C for 30 min before crystallisation. Crystals of UgpB in complex with GPC were grown by sitting-drop vapour diffusion in 96-well plates (Swiss-Ci) using a Mosquito liquid handling system (TTP LabTech) by mixing 1:1, 2:1 and 1:2 volumes (150 nL) of concentrated methylated UgpB (7 mg/mL) with reservoir solution. UgpB crystals typically grew within three days at 22 °C in 0.2 M $MgCl_2$, 0.1 M tris pH 8.5, 20 % w/v PEG 8,000. The UgpB crystals were cryoprotected with 20 % glycerol and flash frozen in liquid N_2 prior to data collection.

The X-ray diffraction data for the ligand bound UgpB crystals were collected at the I04-1 beamline of Diamond Light Source. The diffraction data were indexed, integrated and scaled with XDS (Kabsch et al,. 2010) through the XIA2 pipeline and CCP4 suite of programmes (Collaborative computational project. 2019). Initial phases were determined by molecular replacement using PHASER (McCoy et al,. 2007) and the separate domains (Domain I residues 34-153 and 305-378/Domain II residues 154-304 and 379-424) of the apo-UgpB structure as two ensembles as a search model (PDB 4MFI) specifying to search for 4 copies in the asymmetric unit. Autobuild (Terwilliger et al,. 2008) was initially used for model building followed by iterative cycles of alternating manual rebuilding in COOT (Emsley et al,. 2004) and reciprocal space crystallographic refinement with PHENIX-REFINE (Afonine et al,. 2012) assigning each domain as a

separate TLS group. The coordinates for the GPC ligand were downloaded from the PDB ligand expo (PDB CH5) and fitted into unoccupied electron densities in all four chains of the asymmetric unit. The restraints for use in refinement were calculated using REEL (Moriarty et al,. 2017). Magnesium ions and glycerol molecules were also fitted into the unoccupied densities as well as waters. Methylated lysine (MLZ) was fitted at position 161 in all chains.

The model of the ligand-bound structure comprises residues 36-428 in all chains (A-D), with an additional 1 residue in chains B and D and 2 residues in chain C. There is a disordered region between residues 355-366 in chains C and D and these residues were not modelled. No Ramachandran outliers were identified and structure validations were done by MolProbity (Chen et al,. 2010). Figures were prepared using Pymol (The PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC), except for those showing electron density, which were prepared using CCP4mg (McNicholas et al,. 2011).

2.6.3 Crystallisation and structure determination GPS bound *Mtb* UgpB

For co-crystallisation experiments methylated *Mtb* UgpB was incubated with 20 mM glycerophosphserine (GPS) and incubated at 4 °C for 30 min before crystallisation. Crystals of UgpB in complex with GPS were grown by sitting-drop vapour diffusion in 96-well plates (Swiss-Ci) using a Mosquito liquid handling system (TTP LabTech) by mixing 1:1, 1:2 and 2:1 ratiometric volumes of protein with reservoir solution to produce drop sizes of 0.3-0.45 μL. The crystallisation conditions that were initially screened included JCSG-*plus*, Morpheus, Midas and Structure Screen I and II (Molecular Dimensions). We got crystals in condition D4 from JCSG-*plus* (0.1 M sodium acetate, 0.2 M lithium sulphate, pH 4.5, 30 % w/v PEG8000) and therefore we optimised using this condition. The Hampton additive screen (Hampton Research) was used at manufacturers guidance to give a final 10 % additive concentration In the additive screen several crystals were identified after 2 weeks in the additives: ethylene glycol (3.0 %), acetone (4.0 %), ethyl acetate (0.5 %) and trimethylamine hydrochloride (0.01 M). The crystals were flash frozen in liquid nitrogen prior to diffraction data collection at the I-24 beamline (Diamond Light Source). No cryo-protection was added due to the high concentration of PEG8000 (30 % w/v) in the crystallisation condition. The crystal that produced high-resolution diffraction data (1.76 Å) grew in the additive ethylene glycol (3.0 %).

To solve the phases molecular replacement (PHASER) (McCoy et al., 2007) was used specifying the most well refined chain A of GPC bound *Mtb* UgpB as the single search model. The calculated phases were well defined and the model was visualised (COOT) (Emsley et al., 2004) and appeared to be identical to GPC bound *Mtb* UgpB. However the specific regions of amino acids, 90-93 did not fit well to the density and had to be manually adjusted. No methylated lysines were identified in the structure despite reductively methylating the protein prior to crystal trials. At the centre of the structure between the two-α/β domains an electron dense region was identified as the putatively bound GPS ligand. The glycerol and phosphate electron densities were clearly defined and GPS (PDB ID GSE) was manually fitted and the molecular restraints were calculated (eLBOW) (Adams et al., 2010). To finally build the model, H_2O molecules were fitted into the appropriate densities.

All STD-NMR experiments were performed by: Dr Ridvan Nepravishta and Dr Jesus Angulo, School of Pharmacy, University of East Anglia.

2.6.4 ^{1}H STD-NMR experiments

Mtb UgpB was expressed and purified as described (2.4.1 and 2.4.6). The protein was buffer exchanged into deuterated PBS pH 7.5 by ultrafiltration (Vivaspin 10 kDa MWCO) by 5x 10 mL additions of deuterated PBS and concentrating, bringing the volume from 10 mL to 0.5 mL after each exchange. After buffer exchange, *Mtb* UgpB was concentrated to 3.1 mg/mL. For all STD-NMR experiments the ligands were dissolved in deuterated PBS. For the complex of *Mtb* UgpB/GPC the protein concentration was 68 μM and the ligand concentration was 5 mM. STD-NMR spectra were acquired on a Bruker Avance 500.13 MHz at 298 K. The on- and off-resonance spectra were acquired using a train of 50 ms Gaussian selective saturation pulses using a variable saturation time from 0.5 s to 4 s, and a relaxation delay (D1) of 4 s. The water signal was suppressed using the watergate technique while the residual protein resonances were filtered using a $T_{1\rho}$-filter of 50 ms. All the spectra were acquired with a spectral width of 5 kHz and 32K data points using 128 scans. The on-resonance spectra were acquired by saturating at 0.77 or 6.78 ppm while the off-resonance spectra were acquired by saturating at 40 ppm. Instead, for the *Mtb* UgpB/GPIP4 complex, the protein concentration was 35 μM while the ligand concentration was 2.5 mM. STD-NMR spectra were acquired on a Bruker Avance 800.23 MHz at 278 K. The on- and off-resonance spectra were acquired using a train of 50 ms Gaussian selective saturation pulses using a variable saturation time from 0.5 s to 4 s and a relaxation delay (D1) of 5 s. The water signal was suppressed by using the excitation sculpting technique while the residual protein resonances were filtered

using a $T_{1\rho}$ -filter of 24 ms. All the spectra were acquired with a spectral width of 12.82 kHz and 32K data points using 64 scans. The on-resonance spectra were acquired by saturating at 0.7 or 6.67 ppm while the off-resonance spectra were acquired by saturating at 40 ppm. To get accurate structural information from the STD-NMR data and in order to minimize the T_1 relaxation bias, the STD build up curves were fitted to the equation $STD(t_{sat}) = STD_{max}*(1-exp(-k_{sat}*t_{sat})$ calculating the initial growth rate STD_0 factor as $STD_{max}*k_{sat} = STD_0$ and then normalizing all of them to the highest value.

All CORCEMA-ST calculations were performed by: Dr Ridvan Nepravishta and Dr Jesus Angulo, School of Pharmacy, University of East Anglia.

2.6.5 CORCEMA-ST calculations

The CORCEMA-ST software was used to calculate the theoretical STD intensities from the crystallographic structure of the *Mtb* UgpB/GPC complex. The parameters used for the calculations were: saturation frequency range 0-1.1 ppm; protein correlation time 45 ns; K_d 0.005 mM; order parameter 0.85; ligand correlation time 0.3 ns; ρ-leak 0.35 s; τ_m 10 ps; cutoff 8 Å; $[L]_0$ 5 mM; $[E]_0$ 68 μM; field 500 MHz. While for the *Mtb* UgpB/GPIP4 complex model obtained from docking calculations, the theoretical STD intensities were calculated using the following parameters: saturation frequency range 0-0.9 ppm; protein correlation time 45 ns; K_d 1 mM; order parameter 0.85; ligand correlation time 0.3 ns; ρ-leak 0.1 s; τ_m 10 ps; cutoff 8 Å; $[L]_0$ 5 mM; $[E]_0$ 68 μM; field 800 MHz. The calculations were repeated in order to have the best fitting possible between the calculated and the experimental ^{1}H STD-NMR intensities. For CH_2 protons showing the same chemical shift an averaged calculated ^{1}H STD-NMR intensity was assumed. NOE factor was used to evaluate the best fit to the experimental data.

All AutoDock Vina docking calculations were performed by: Dr Ridvan Nepravishta and Dr Jesus Angulo, School of Pharmacy, University of East Anglia.

2.6.6 AutoDock Vina docking calculations

Autodock tools were used to prepare for docking both the ligand GPIP4 and the *Mtb* UgpB protein. The calculations were performed by positioning a grid of 20 x 24 x 22 Å in the center of the binding site of *Mtb* UgpB, which was maintained rigid while the ligand was considered flexible. The calculations were performed using Autodock Vina.

All DEEP STD-NMR experiments were performed by: Ridvan Nepravishta and Jesus Angulo, School of Pharmacy, University of East Anglia.

2.6.7 DEEP STD-NMR

DEEP-STD factors were obtained. Briefly, frequencies derived from shiftx2 for aliphatic and aromatic residues present in the binding site of *Mtb* UgpB were used for the position of the saturating selective pulse. The STD obtained using a saturation of 0.5 s on aliphatic or aromatic regions (0.7 or 6.67 ppm, respectively) were then used to calculate the DEEP-STD factors.

2.7 Chemistry

2.7.1 General methods

Unless otherwise stated, chemicals and solvents used in the synthesis were purchased from Sigma Aldrich and used as supplied without further purification. Methanol (MeOH), dichloromethane (DCM), chloroform, pyridine, triethylamine, ammonium hydroxide and magnesium sulphate ($MgSO_4$) were purchased from Fisher Scientific at laboratory reagent grade. Anhydrous tetrahydrofuran (THF) 99.9 %, anhydrous dichloromethane (DCM) 99.8 %, deuterium oxide (D_2O) 99.9 % was purchased from Sigma Aldrich. Deuterochloroform ($CDCl_3$) 99.8 % was purchased from Apollo Scientific. All reactions were performed using oven-dried glassware and heat transfer was achieved using an oil bath.

Thin layer chromatography (TLC) was carried out using aluminium-backed sheets coated with 60 F_{254} silica gel (Merck). Visualization was carried out using the stains (Table 2.10). Flash column chromatography was carried out using silica gel 60, 35-75 µm (Merck) as the stationary phase.

Chemistry procedures and analyses were carried out with the assistance of Dr Collette Guy (Professor Matthew Gibson's Laboratory, Chemistry Department, University of Warwick).

2.7.2 Nuclear magnetic resonance spectroscopy

Proton (^1H), carbon (^{13}C) and phosphorous (^{31}P) NMR spectra were obtained at 298 K. All NMR spectra were recorded on either the Bruker DPX-300 or DPX-400 instruments and spectra were fully assigned using COSY, HSQC and HMBC. Chemical shifts were stated in parts per million (ppm) using the residual solvent as internal standard. Coupling constants (*J*) are reported in hertz (Hz) using the abbreviations: s singlet, d doublet, t triplet, q quartet, quin quintet, m multiplet, br broad.

2.7.3 Mass spectrometry

Low resolution mass spectra were recorded on a Bruker Esquire 2000 instrument using electrospray ionisation (ESI), negative/positive mode where indicated. High-resolution mass spectra were recorded on a Bruker HCT instrument. M/z values are reported in Daltons.

2.7.4 Chemoenzymatic production of glycerophosphoethanolamine (GPE), glycerophosphoserine (GPS) and glycerol-3-phosphocholine enantiomer (GPC):

Phospholipase A1 from *Aspergillus oryzae* (Sigma Aldrich) was dialysed into PBS (Table 2.1) overnight at 4°C prior to use to remove sorbitol. The enzymatic reaction contained either 50 mg of 1,2-Dipalmitoyl-*sn*-glycero-3-phosphoethanolamine (Sigma Aldrich), 25 mg of 1,2-Diacyl-*sn*-glycero-3-phospho-L-serine (Sigma Aldrich) or 12.5 mg of 2,3-Dipalmitoyl-*sn*-glycerol-1-phosphocholine (Sigma Aldrich) in an organic-aqueous media of 800 μL hexane and 138 μL H_2O (5.8:1 v/v) and was heated at 50°C for 10 min prior to the addition of Phospholipase A1 from *Aspergillus oryzae* (1 μL Phospholipase A_1 per 1 mg of phospholipid). The reaction mixtures were heated at 50 °C and stirred at 300 rpm for 48 h. The solvent was evaporated under vacuum and the reaction mixture was dissolved in 5 mL of water and extracted with chloroform 3 x 25 mL. The aqueous phase was separated and the phospholipase A_1 enzyme removed by ultrafiltration (Amicon, 10 kDa molecular weight cut off). The collected filtrate was freeze-dried.

2.7.5 Characterisation of GPE, GPS and GPC enantiomer by NMR:

GPE: ^{1}H NMR (400MHz, D_2O) δ_{ppm} 3.99 – 4.08 (2H, m, POC$\underline{H}_2$C$\underline{H}_2$N), 3.75- 3.91 (3H, m, POC$\underline{H}_2$C$\underline{H}$C$\underline{H}_2$), 3.49 – 3.63 (2H, m, POCH$_2$CHC$\underline{H}_2$), 3.20 (2H, t, *J* = 5.0 Hz, POCH$_2$C$\underline{H}_2$N). ^{13}C NMR (100MHz, D_2O) δ_{ppm} 70.7 (C$\underline{H}$), 66.5 (OC$\underline{H}_2$), 62.0 (OC$\underline{H}_2$), 61.8 (OC$\underline{H}_2$), 40.0 (NC$\underline{H}_2$). ^{31}P NMR (161 MHz, D_2O) δ_{ppm} 0.42.

GPS: ^{1}H NMR (400MHz, D_2O) δ_{ppm} 4.17 – 4.26 (2H, m, POC$\underline{H}_2$CHN), 4.01 – 4.08 (1H, m, POCH$_2$C$\underline{H}$N), 3.71 – 3.93 (3H, m, 3H, m, POC$\underline{H}_2$C$\underline{H}$C$\underline{H}_2$), 3.48-3.65 (2H, m, POCH$_2$CHC$\underline{H}_2$). ^{13}C NMR (100MHz, D_2O) δ_{ppm} 178.5 (C$\underline{=}$O), 70.6 (CH$_2$C$\underline{H}$(OH)CH$_2$), 66.5 (OC$\underline{H}_2$), 63.7 (POC$\underline{H}_2$CHN), 62.0 (OC$\underline{H}_2$), 54.4 (POCH$_2$C$\underline{H}$N). ^{31}P NMR (161 MHz, D_2O) δ_{ppm} 0.08.

GPC enantiomer: ^{1}H NMR (400MHz, D_2O) δ_{ppm} 4.18 – 4.29 (2H, m, POC$\underline{H}_2$C$\underline{H}_2$N), 3.74 – 3.90 (3H, m, POC$\underline{H}_2$C$\underline{H}$C$\underline{H}_2$), 3.46 – 3.63 (4H, m, POCH$_2$CHC$\underline{H}_2$ + POCH$_2$C$\underline{H}_2$N), 3.14 (9H, s, C$\underline{H}_3$). ^{13}C NMR (100MHz, D_2O) δ_{ppm} 70.6 (CH$_2$C$\underline{H}$(OH)CH$_2$), 66.5 (OC$\underline{H}_2$), 66.0 (OC$\underline{H}_2$), 62.0 (OC$\underline{H}_2$), 59.4 (NC$\underline{H}_2$), 53.9 (C$\underline{H}_3$).

Chapter 3: Cloning, expression and purification of *Mycobacterium tuberculosis* UgpB

3.1 Introduction

3.1.1 Introduction

The expression and purification of soluble, correctly folded and physiologically active *Mycobacterium tuberculosis* (*Mtb)* proteins is a major challenge and a requirement to elucidate their biochemical structural properties. The bacterium *Escherichia coli* is the most widely used host for recombinant protein expression because it is well-studied, amenable to genetic methods and can be easily cultured (Rosano et al., 2014). The most widely used *E. coli* expression system uses the T7 RNA polymerase from the T7 bacteriophage and the T7 promoter (Tabor et al., 1985). In the commonly used expression strain BL21 (DE3), the T7 RNA polymerase gene is inserted into the bacterial chromosome under the control of the inducible *lacUV5* operon (Siebenlist et al., 1980). The gene of interest is cloned into a plasmid under the control of the T7 promoter that can only be transcribed by the T7 RNA polymerase and is transformed into the bacteria. The lac repressor (*lacI*) binds to the operator sites (*lacO*) in the promoter of the *lacUV5* operon and prevents the T7 RNA polymerase sequence from being transcribed. However, addition of the inducer isopropyl-β-D-1-thiogalactopyranoside (IPTG), an analogue of allolactose that cannot be metabolised, releases *lacI* from the operator enabling the T7 RNA polymerase to be transcribed and to transcribe the gene of interest.

However despite the advantages of *E. coli* for protein expression it is not always the most appropriate host to express mycobacterial proteins as they can be toxic (Melief et al., 2018) and have a tendency to form insoluble inclusion bodies (Baneyx, 1999). The molecular basis of the difficulties of expressing *Mtb* proteins in *E. coli* are not fully understood, however potential contributing factors are due to differences in the guanidine + cytosine (G+C) content of these organisms genomes. The G+C content of *Mtb* is 65.6 % (Cole et al., 1998) and for *E. coli* is 50.8 % (Blattner et al., 1997). Large differences in the G+C content of genes likely results in different codon usage and requirements for different cellular machinery to facilitate normal transcription and translation of the *Mtb* proteins (Bellinzoni et al., 2003). Additional important factors may be involved for example *Mtb* specific chaperones may be required for the correct folding of *Mtb* proteins and maintaining protein stability (Goldstone et al., 2008). Also differences in the composition of the mycobacterial cell membrane and cell wall may be

particularly important to successfully express *Mtb* membrane proteins. (Korepanova et al., 2009)

3.1.2 Alternative host expression systems to *E. coli*

Alternative host expression systems to *E. coli* have been developed to successfully express and purify *Mtb* proteins. These different host expression systems are diverse ranging from eukaryotic virus-infected-insect cells, yeast, bacterial Gram-positive and Gram-negative species and the closely related actinobacteria and mycobacterial species (Atkins et al., 1994) (Gurvitz et al., 2009) (Kieser et al., 1986) (Goldstone et al., 2008).

A baculovirus-insect cell eukaryote recombinant gene expression system was used to successfully express *Mtb* chaperonin 10 (Cpn1) and despite the high G+C content of the gene (64 %) the protein was purified to a high yield (Atkins et al., 1994). Other studies have also used similar systems to express *Mtb* NAD^+ synthase (NadE) (Bellinzoni et al., 2002) and *Mycobacterium bovis* antigen P36 (Bigi et al., 1999). However this system is not routinely used because of practicalities associated with setting up an insect cell expression laboratory. Also non-physiological post-translational modifications can be introduced into the expressed proteins because a prokaryote gene is being expressed in a eukaryote organism. Another eukaryote expression system that has been used to express *Mtb* proteins and investigate physiological function is *Saccharomyces cerevisiae* (Gurvitz et al., 2009). However this system can be disadvantageous because yeast has a longer growth time, doubling every 1.5-2 hours at 30 °C compared to every 20-30 minutes for *E. coli* and they can produce hyper-glycosylated proteins and introduce non-physiological post-translational modifications. *Pseudomonas putida* has a high G+C content genome of 61.5 % similar to the *Mtb* genome and its metabolism is versatile and tolerant to chemical agents and oxidative stress (Ramos-Gonzalez et al., 2005). *Mtb* cytosolic carbon-carbon hydrolase HsaD was successfully expressed in this system and its structure determined by X-ray crystallography (Lack et al., 2008). However this system is not routinely used because the vectors for the production of protein are not as well established.

A potentially better approach is to express *Mtb* proteins in a more closely related bacterial species (Bashiri et al., 2015). For example *Streptomyces* are part of the actinobacteria family, and they also have a high G+C (>70 %) content similar to mycobacteria (Romero-Rodriguez et al., 2015). A study identified that *Streptomyces lividans* could efficiently use a high proportion of mycobacterial promoters and

translational signals suggesting that it could be adapted to express *Mtb* proteins (Kieser et al., 1986) and since this study the 45/47 kDa *Mtb* APA protein (Rv1860) was successfully expressed (Lammertyn et al., 1998). However in this system recombinant proteins are secreted directly into the culture medium and this can be a problem for producing proteins in the correct conformation at sufficient quantities. Many of these alternative hosts are not as convenient and practical to use compared to *E. coli* however a single solution system to express *Mtb* proteins does not exist and investigations often require trial and error.

3.1.3 *Mycobacteria* host expression systems

Mycobacterium smegmatis is a closely related mycobacterial species sharing approximately 70 % of its genome with *Mtb*. It has the advantage of being non-pathogenic and relatively fast growing as its doubling time is only approximately 4 hours (Snapper et al., 1990). *M. smegmatis* has been widely adapted as a model organism to study *Mtb* and has also been adapted to express *Mtb* proteins (Goldstone et al., 2008). Its high G+C genome content and shared genes likely result in it having similar codon usage and myocbacterial specific heat shock, stress-induced and chaperones likely aid the correct folding, solubility and stability of expressed proteins (Poulsen et al., 2010). These effects may be further enhanced by mycobacterial specific metabolites and ligand cofactors as well as binding proteins (Bashiri et al., 2012). Its cell membrane and cell wall also have a similar lipid composition that may be necessary to express *Mtb* membrane proteins (Etienne et al., 2005).

To date 84 X-ray crystal structures of *Mtb* proteins expressed in *M. smegmatis* are available in the protein data bank (RSCB PDB 2019). However despite these advantages the yields of *Mtb* proteins expressed in *M. smegmatis* vary and sometimes result in slow cell growth (Saida et al., 2006). However the slower growth rate of *M. smegmatis* may be important for correct protein folding and for producing potentially toxic proteins. *M. smegmatis* mc^2155 is available for efficient plasmid uptake (Snapper et al., 1990). A modified strain of mc^2155 called *M. smegmatis* mc^24517 is now available for protein expression. *M. smegmatis* mc^24517 also expresses T7 RNA polymerase under the control of the acetamidase inducible promoter, in the vector pYUB1232 (Wang et al., 2010). *M. smegmatis* mc^24517 is used to routinely express genes of interest in vectors controlled by the T7 promoter (Wang et al., 2010). A related mycobaterial species called *Mycobacterium vaccae* has also been utilised to successfully express and purify the tetrameric iron-dependent superoxide dismutase of *Mtb* and determine its structure by X-ray crystallography (J. B. Cooper et al., 1995). However *M. vaccae* is not routinely used because most mycobacterial vectors contain

a kanamycin resistance gene and some species of *M. vaccae* do not efficiently express the aminoglycoside phosphor-transferase genes involved in kanamycin resistance (Medeiros et al., 2002). As a result *M. smegmatis* has been routinely developed and used as a strain to produce *Mtb* proteins.

3.1.4 Routes to optimise protein production

Soluble protein expression and purification are major hurdles however genetically encoded tags are available to assist these processes. The poly-histidine tag is a very small tag consisting of a repeating sequence of 6-10 histidine residues. The poly-histidine tag is easily incorporated into a plasmid at either the N or C-terminus of the protein of interest and when expressed has a high affinity for divalent metal cations (Bornhorst et al., 2000). The glutathione S-transferase (GST) tag is a larger 26 kDa protein tag inserted at the N-terminus of the protein of interest to enhance protein solubility and stability (Schafer et al., 2015). However, due to the large size of the GST tag it may affect protein crystallisation and require proteolytic cleavage. The small unbiquitin-like modifier (SUMO) tag is an 11 kDa protein from *Saccaromyces cerevisiae* Smt3 (Saitoh et al., 1997). Similar to the GST tag if SUMO is fused to the N-terminus of a protein of interest it can result in enhanced soluble protein expression levels. However due to its large size it may interrupt normal protein function and therefore must be cleaved. The green fluorescent protein (GFP) is a 27 kDa protein and is used primarily as a fluorescent reporter. If GFP is inserted at the C-terminus of a gene of interest, GFP fluorescence can be used to monitor expression of correctly folded protein. GFP has also been successfully adapted as a reporter for *Mtb* protein expression in *M. smegmatis* (Radhakrishnan et al., 2018).

3.1.5 What we know about *Mtb* UgpB protein expression

Mtb ugpB (Rv2833c) encodes for the substrate binding protein *Mtb* UgpB and is located in a putative four-gene *ugpABCE* ATP-binding cassette (ABC) transporter operon. A previous study (Jiang et al., 2014) produced recombinant expressed UgpB as a N-terminal GST fusion protein in *E. coli*. In these studies the structure of UgpB in the ligand-unbound conformation was solved by X-ray crystallography to 1.5 Å and ligand binding to glycerol-3-phosphocholine (GPC) was identified by isothermal calorimetry (Jiang et al., 2014). However the diversity of the ligands tested was limited and the inability to solve a ligand bound structure prompted us to carry out a more extensive biochemical analysis starting by expressing Mtb UgpB.

3.2 Hypotheses and Aims

The primary hypothesis of this chapter was that an N-terminal truncated *Mtb* UgpB substrate-binding protein could be expressed using the pYUB1062 vector in *M. smegmatis* $mc^2$4517. The protein could be purified to homogeneity using a combination of immobilised metal affinity, ion exchange and gel filtration chromatography, and its identity confirmed by mass spectrometry. An additional hypothesis was that alanine site directed mutants of the protein could be produced using the same method, and that the secondary structure profiles of these mutants would be similar to the apo protein. To test this hypothesis circular dichroism was used.

The specific aims of the work presented in this chapter were:

- To express and purify *Mtb* UgpB, the substrate-binding protein of the UgpABCE ATP-binding cassette transporter.
- To clone, express and purify site-directed mutants of *Mtb* UgpB.

3.3 Results

3.3.1 Identification of UgpB in *Mycobacteria*

The *ugpB* gene is found in *Mtb* and *Mycobacterium bovis* the causative agents of tuberculosis in humans and primarily bovine mammals and shares 99.3 % amino acid sequence (NCBI BLAST). However in *Mycobacterium bovis* BCG the vaccine strain for *Mtb* (Calmette, 1922), the *ugpB* gene is frame shifted at nucleotide position 631 of the *ugpB* gene, where cytosine is replaced by thymine resulting in the formation of a TAG stop codon that truncates the translated sequence (Brosch et al., 2007). In contrast, from comparative genomics (Mycobrowser, 2019), the *ugpB* gene is not present in *M. marinum* the causative agent of tuberculosis in fish or *M. leprae* the causative agent of leprosy that is considered to have a core genome that is massively smaller compared to the *Mtb* genome (Vissa et al., 2001). Interestingly the *ugpB* gene is only found in pathogenic mycobacteria species whereas it is not found in the closely related fast growing and non-pathogenic *M. smegmatis* species therefore these differences may reflect the different physiological environments of these organisms.

3.3.2 Operon organisation of *Mycobacterium tuberculosis* UgpABCE transporter

The UgpABCE transporter is organised in a putative 4-gene operon (Figure 3.1). The *ugpC* gene is 1083 bp nucleotides and encodes the predicted nucleotide binding protein and is approximately 39.3 kDa. The *ugpB* gene is 1311 bp nucleotides and encodes the substrate binding protein and is approximately 46.7 kDa. The *ugpA* and *ugpE* genes are 912 bp nucleotides and 828 bp nucleotides and encode the membrane spanning proteins and are approximately 34.1 kDa and 30.7 kDa. Adjacent to the core transporter genes is echA16, a probable enoyl-CoA hydratase and dinF a possible DNA damage inducible protein. However, the genetic annotation of these surrounding genes did not indicate the role of the UgpABCE transporter.

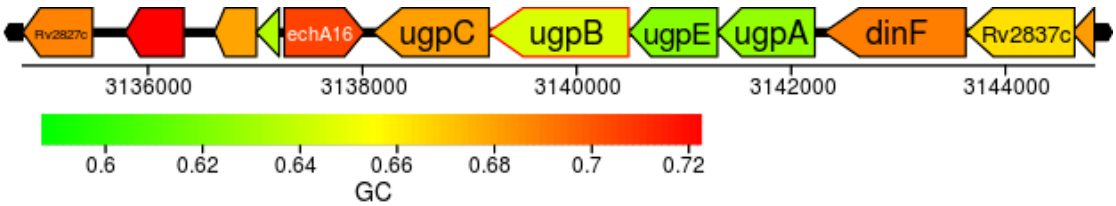


Figure 3.1: Operon organisation of *Mycobacterium tuberculosis* Ugp transporter. Genes are colour coded according to G+C content.

3.3.3 Comparison with *E. coli* UgpB

Multiple sequence alignment of *Mtb* UgpB and *E. coli* UgpB, which are the only known structures of UgpB (PDB ID 4MFI) (PDB ID 4AQ4) (Jiang et al., 2014) (Wuttge et al., 2012) (Figure 3.2) revealed large differences at the amino acid level with only 26.4 % similarity over 83 % read. Based on these differences it is therefore possible that these proteins have evolved to carry out different physiological functions. Interestingly, the *Mtb* UgpB protein encodes a lipobox motif [Lys (Ala/Ser) (Gly/Ala) Cys] at the N-terminus. The cysteine residue in this motif may enable the protein to be lapidated and attach to the cell membrane. In direct contrast, *E. coli* UgpB does not encode a lipobox motif at the N-terminus therefore the protein is unlikely to be lipidated. These differences may reflect the different cell wall structures of these organisms. The Gram –ve *E. coli* bacteria has an inner and outer membrane separated by the periplasm. The absence of a lipobox motif in *E. coli* UgpB suggests that the protein exists freely in the periplasm and is not bound to the lipid membrane. Conversely, *Mtb* is a pseudo Gram +ve bacteria and does not have a periplasm therefore the encoded lipobox motif may be required for the protein to be attached to the cell membrane or lipids found in the cell wall. This potential ability to attach to the cell wall may provide a mechanism through which the protein can acquire nutrients external to the cell wall. I have only attempted to express and purify the N-terminal truncated protein because it is likely to be more soluble and therefore more amenable to biochemical characterisation. However, this N-terminal region may be intergral for the transporters function within the cell.

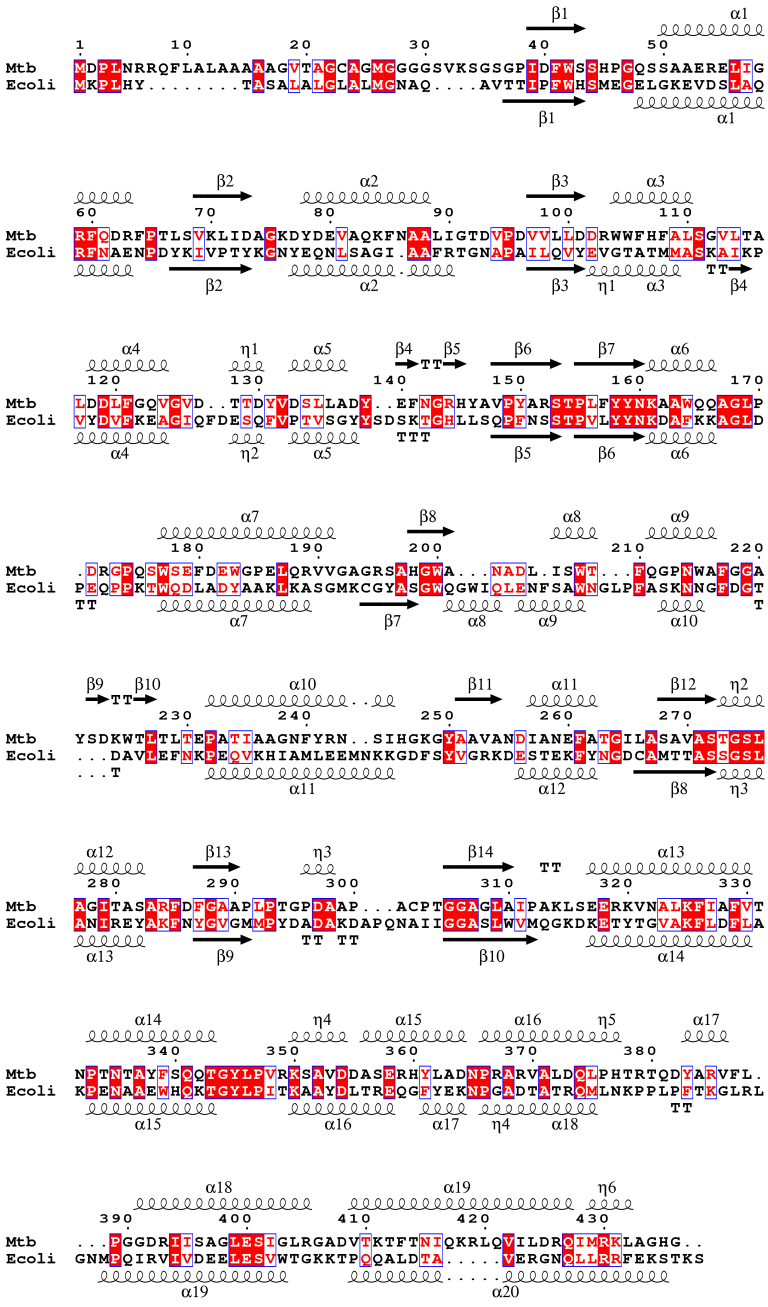


Figure 3.2: Sequence alignment of UgpB from *Mycobacterium tuberculosis* with UgpB from *Escherichia coli*. The sequence alignment was generated using Clustal Omega and ESPript version 3. Identical residues are indicated by a red background and conserved residues by red characters. The secondary structure elements of *Mtb* UgpB are shown above the sequences and the secondary structure elements of *E. coli* UgpB (4AQ4) are shown below the sequences.

3.3.4 Cloning of *ugpB*-pYUB1062

The *Mtb ugpB* (Rv2833c) gene was previously expressed with a GST tag (Jiang et al., 2014) however attempts in our lab to replicate these results were unsuccessful. Various systems were tried within the lab including expressing the gene in

Pseudomonas putida (Lack et al., 2008). Eventually *Mtb* UgpB was successfully expressed in *M. smegmatis* mc^{2}4517.

The *Mtb ugpB* (Rv2833c) gene was amplified and cloned into the vector pYUB1062 with a C-terminal hexa-His tag (Figure 3.3) by Dr James Harrison, as described (Chapter 2, 2.3.3.1). In this construct the *Mtb* UgpB protein was truncated at the N-terminus where residues 1-34 were removed. Residues 1-34 are predicted by TMpred (ExPASy) to form a single trans-membrane anchor helix and secreted signal sequence and lipobox motif that may decrease protein expression and solubility (Cuthbertson et al., 2005). Removal of the first 34 amino acids containing this lipobox motif was found to enhance soluble protein expression in *M. smegmatis*.

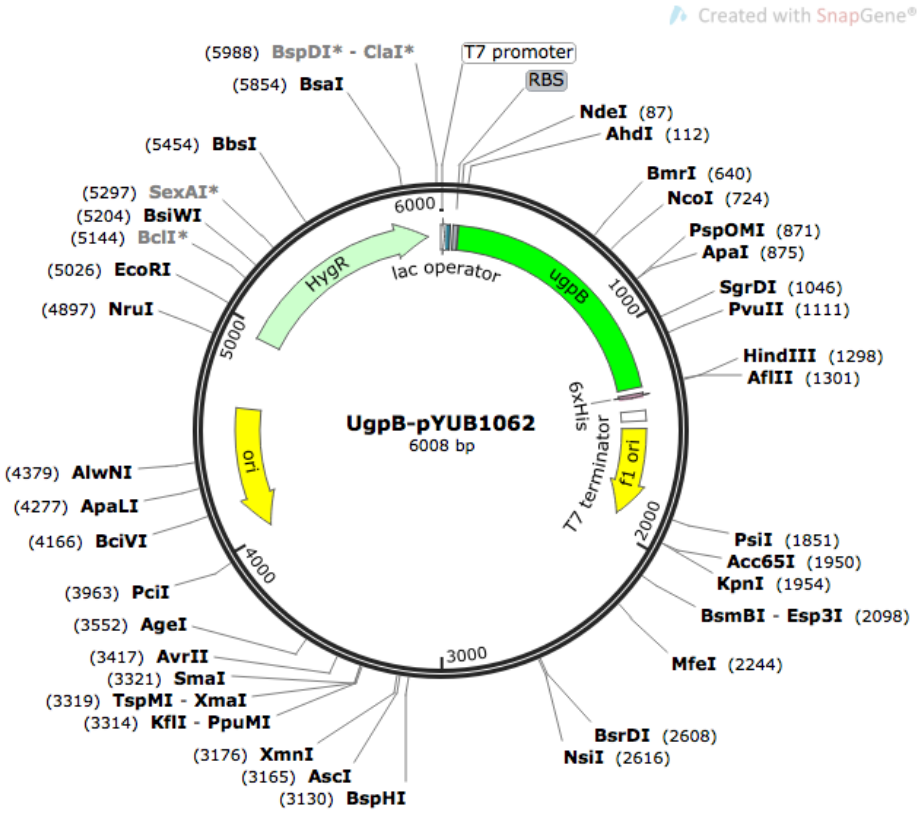


Figure 3.3: *Mtb* UgpB-pYUB1062 plasmid map. Abbreviations: HygR hygromycin resistance gene, f1 ori and ori origins of replication in *E. coli* and *Mycobacteria*.

3.3.5 Expression and purification of UgpB-pYUB1062

The truncated *Mtb ugpB*-pYUB1062 construct was electroporated into *Mycobacterium smegmatis* mc^{2}4517 electro-competent cells. The cells were incubated in 6 L of LBGT media at 37 °C until optical density 0.4 was reached. The cells were then induced by the addition of 0.2 % acetamide and expression was continued for 20 hours at 37 °C. This expression condition resulted in protein being produced and was used as the optimal condition for subsequent expressions and purifications.

After expression the cells were suspended in UgpB lysis buffer (500 mM NaCl, 20 mM NaH_2PO_4, 0.1 % (v/v) triton X-100, pH 7.4) (Chapter 2, Table 2.1) and lysed by sonication. The cell lysate was passed over a Co^{2+} immobilised metal affinity chromatography (IMAC) column. Following immobilisation of *Mtb* UgpB to the Co^{2+} column, the protein was purified through a series of wash-steps and eluted with a step-wise imidazole gradient. The fractions were collected and analysed by SDS-PAGE (Figure 3.4A). The expected *Mtb* UgpB protein size is approximately 44 kDa and the gel indicated a large amount of correctly expressed protein in the 50-1000 mM imidazole fractions (Figure 3.4A). A high intensity protein band was also observed at approximately 65 kDa that is most likely to be the chaperone GroEL. The imidazole elution fractions (50-1000 mM) contained the largest amount of *Mtb* UgpB protein however it was not completely pure at this stage therefore ion-exchange chromatography was carried out. First, the 50-1000 mM imidazole fractions were combined and dialysed (150 mM NaCl, 25 mM HEPES, 10.0 % glycerol pH 7). The dialysed protein was passed over a QHP anion exchange chromatography column to separate the proteins based on overall charge. The theoretical pI of the *Mtb* UgpB protein is 5.64 (ExPASy ProtParam) (Wilkins et al., 1999) so the protein should have an overall negative charge and, in theory, would not bind to the negatively charged anion-exchange column. Under these purification conditions the *Mtb* UgpB protein did not bind to the QHP column and, instead, was collected in the flow through and 150 mM NaCl wash (Figure 3.4B). The 70 kDa and 35 kDa contaminants observed from the first Co^{2+} column appeared to bind to the QHP column and were eluted with an increased concentration gradient of NaCl however, the 15 kDa and 11 kDa contaminants co-eluted with the *Mtb* UgpB protein (Figure 3.4B). The flow through and 150 mM NaCl fractions were collected and concentrated by ultrafiltration (10 kDa cut-off, Amicon Ultra) to approximately 2 mL and passed over a Superdex 200 gel filtration chromatography column to separate the proteins based on size (150 mM NaCl, 25 mM HEPES, 10.0 % glycerol pH 7). 1 mL fractions were collected, analysed by SDS-PAGE (Figure 3.4C) and *Mtb* UgpB eluted (83-103 mL) as a monomer at the expected size of 44 kDa (Figure 3.4D). However the protein was still not completely pure as the 15 kDa and 11 kDa bands remained (Figure 3.4C) suggesting that they could be interacting with *Mtb* UgpB therefore the identities of these proteins were analysed by in-gel trypsin digest mass spectrometry (Figure 3.5).

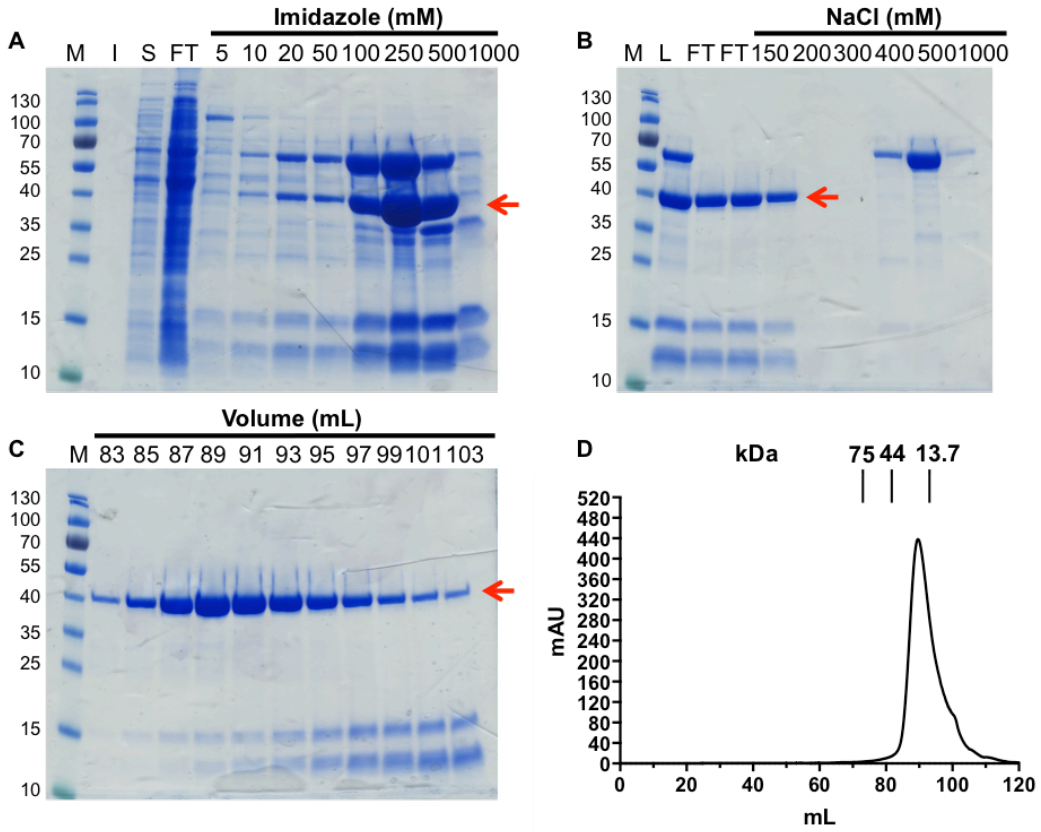


Figure 3.4: SDS-PAGE analysis of the purification of *Mtb* UgpB from *M. smegmatis*. A) Elution of His_6-tagged *Mtb* UgpB from a Co^{2+} IMAC-column. M = molecular weight marker in kDa, IS = insoluble fraction, S = soluble lysate, FT = flow through, numbers 5 – 1000 refer to the imidazole concentration in the elution buffer (units of mM). B) QHP anion exchange chromatography of *Mtb* UgpB following the Co^{2+} IMAC step. L = protein after dialysis, FT = flow through, numbers 150 – 1000 refer to NaCl concentration in the elution buffer (units of mM). C) Size exclusion chromatography of *Mtb* UgpB following anion exchange chromatography with the volumes shown as corresponding to D. D) Size exclusion trace of *Mtb* UgpB.

3.3.6 Analysis of low molecular weight contaminants by mass spectrometry

Both of the contaminant proteins were identified as rplU and rpsK (Figure 3.5) from *M. smegmatis.* The 11 kDa contaminant band was identified as the 50S ribosomal protein rplU, 22 exclusive unique peptides, 33 exclusive unique spectra, 152 total spectra, 101/103 amino acids (598 % coverage). The 15 kDa contaminant band was identified as the 30S ribosomal protein rpsK, 10 exclusive unique peptides, 15 exclusive unique spectra, 41 spectra, 53/138 amino acids (38 % coverage). It is likely that these contaminants, as they are part of the ribosomal translational machinery are a result of expressing proteins in *M. smegmatis* and do not reflect any physiological interactions with *Mtb* UgpB.

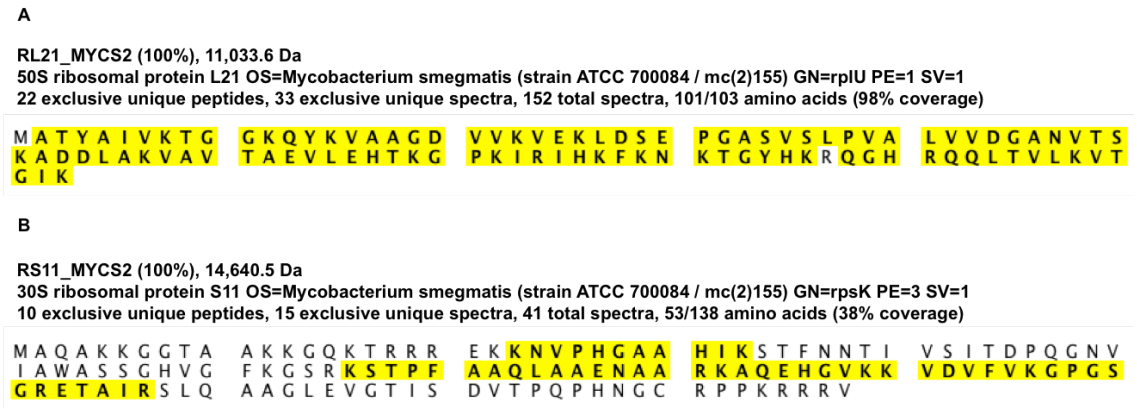


Figure 3.5: In-gel trypsin digest mass spectrometry of low molecular weight contaminants. A refers to the 11 kDa rplU **B** refers to the 15 kDa rpsK. Sequences highlighted in yellow indicate coverage of the peptides identified.

3.3.7 Optimisation of *Mtb* UgpB protein purification

To optimise and increase the purity of *Mtb* UgpB the same expression procedure was carried out (3.3.5) and the purification steps were the same as before apart from the final size exclusion step. Instead a Superdex 75 gel filtration chromatography column was used to improve the separation of the low molecular weight contaminants (Figure 3.6C). A Superdex 75 separates in the molecular weight range (3-70 kDa) whereas Superdex 200 separates in the molecular weight range (10-600 kDa). Following size-exclusion chromatography with the Superdex 75 column in the same buffer (150 mM NaCl, 25 mM HEPES, 10 % glycerol pH 7.0) the fractions were analysed by SDS-PAGE and found to be a higher purity (Figure 3.6C) compared to the Superdex 200 column (Figure 3.4C). No contaminants were observed apart from in the final volume 66-68 mL so this was not pooled with the pure UgpB protein. This separation difference was probably due to differences in the gel filtration chromatography pore size distribution. The fractions containing pure *Mtb* UgpB volumes 51-64 mL (Figure 3.6C) were combined concentrated by ultrafiltration to 2 mg/mL and stored at -80 °C prior to analyses. Therefore this was identified as the best method for optimisation and used in subsequent purifications.

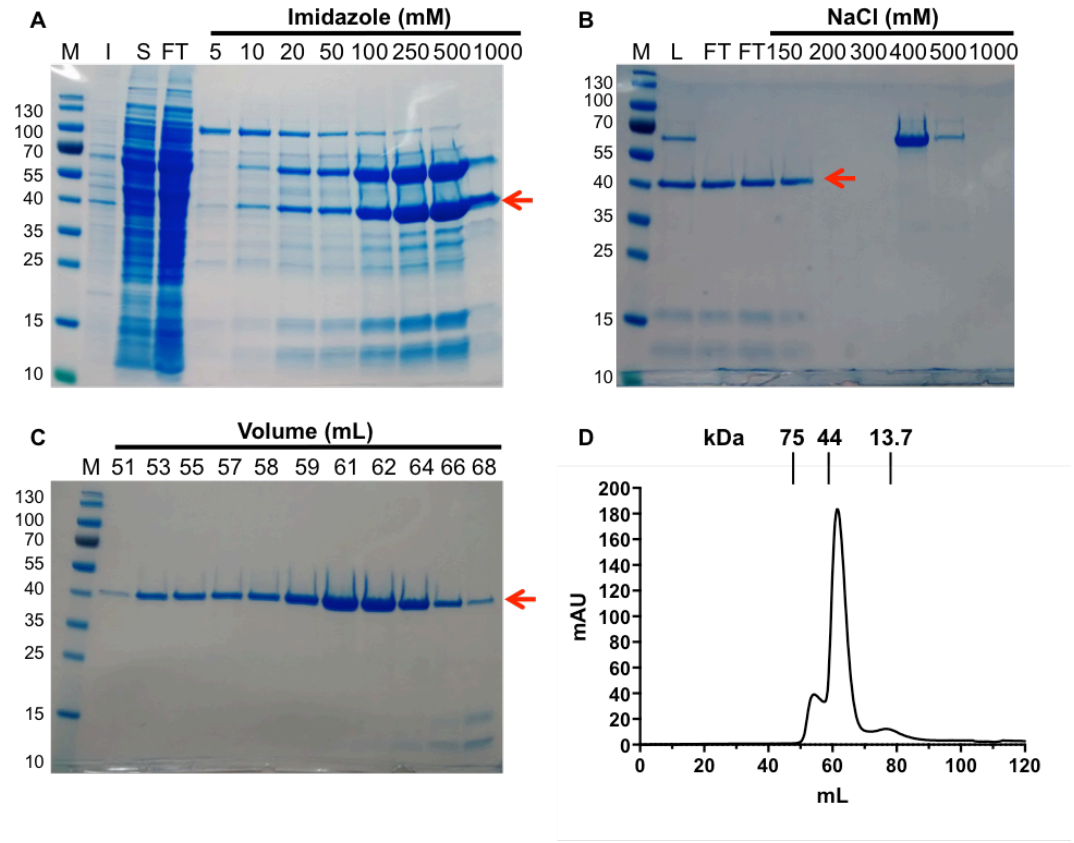


Figure 3.6: SDS-PAGE analysis of the purification of *Mtb* UgpB from *M. smegmatis*. A) Elution of His_6-tagged *Mtb* UgpB from a Co^{2+} IMAC-column. M = molecular weight marker in kDa, IS = insoluble fraction, S = soluble lysate, FT = flow through, numbers 5 – 1000 refer to the imidazole concentration in the elution buffer (units of mM). B) QHP anion exchange chromatography of *Mtb* UgpB following the Co^{2+} IMAC step. L = protein after dialysis, FT = flow through, numbers 150 – 1000 refer to NaCl concentration in the elution buffer (units of mM). C) Size exclusion chromatography of *Mtb* UgpB following anion exchange chromatography with the volumes shown as corresponding to D. D) Size exclusion trace of *Mtb* UgpB.

3.3.8 Mass spectrometry of *Mtb* UgpB

The identity of purified *Mtb* UgpB was confirmed by in-gel trypsin digest mass spectrometry of the approximate 44 kDa SDS-PAGE band (Figure 3.6C). The *Mtb* H37Rv and *M. smegmatis* $mc^2$4517 proteome databases were searched. 25 exclusive unique peptides, 39 exclusive unique spectra, 230 total spectra, 234/436 amino acids (54 % coverage) were identified, confirming the identity of *Mtb* UgpB.

P7169 (100%), 46,726.6 Da
ABC transporter substrate-binding protein OS=Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv) GN=ugpB PE=1 SV=1
25 exclusive unique peptides, 39 exclusive unique spectra, 230 total spectra, 234/436 amino acids (54% coverage)

MDPLNRRQFL AAERELIGRF DDRWWFHFAL ARSTPLFYYN ANADLISWTF AAVANDIANE ACPTGGAGLA SAVDDASERH SIGLRGADVT
ALAAAAAAGVT QDRFPTLSVK SGVLTALDDL KAAWQQAGLP QGPNWAFGGA FATGILASAV IPAKLSEERK YLADNPRARV KTFTNIQKRL
AGCAGMGGGG LIDAGKDYDE FGQVGVDTTD DRGPQSWSEF YSDKWTLTLT ASTGSLAGIT VNALKFIAFV ALDQLPHTRT QVILDRQIMR
SVKSGSGPID VAQKFNAALI YVDSLLADYE DEWGPELQRV EPATIAAGNF ASARFDFGAA TNPTNTAYFS QDYARVFLPG KLAGHG
FWSSHPGQSS GTDVPDVVLL FNGRHYAVPY VGAGRSAHGW YRNSIHGKGY PLPTGPDAAP QQTGYLPVRK GDRIISAGLE

Figure 3.7: In-gel trypsin digest mass spectrometry of *Mtb* UgpB.
Sequences highlighted in yellow indicate coverage of the peptides identified.

3.3.9 Generation of genetically modified derivatives of *Mtb* UgpB

Single site point mutations were introduced into 9 individual amino acid residues based on my structure of *Mtb* UgpB in complex with its GPC ligand and the identification of specific residues involved in ligand binding interactions (Chapter 5). Mutations were introduced by site-directed mutagenesis, replacing each individual residue with alanine using primers designed to incorporate the codon change (Chapter 2, Table 2.4). The primers were designed to have 9-15 bp either side of the codon of interest that was altered to result in a codon change to alanine. The length of the primers was selected to minimise the G+C content to enable efficient extension. The codon sequence for alanine was selected to be either GCC or GCG depending on the most similar surrounding nucleotides and because GCC and GCG have the highest frequency in the *M. smegmatis* mc^2155 proteome (NCBI GenBank) of 45 % and 42 % respectively. GCC and GCG are therefore most likely to successfully result in incorporation of the alanine amino acid. The site-directed mutagenesis used the specific primers (Chapter 2, Table 2.4) and apo-*ugpB*-pYUB1062 as a template plasmid. Following thermocycling the methylated parental plasmid was digested by *dpnI* digest (1 μL, 37 °C, 3 hours), and the purified plasmid was transformed into *E. coli* Top 10 and plated onto LB agar containing hygromycin. For each mutant, 3 colonies were selected and the plasmid was isolated and sequenced with T7 F and T7 R primers (Chapter 2, Table 2.3) to confirm the correct site-directed mutant DNA sequence.

3.3.10 Protein production and purification of *Mtb* UgpB site-directed mutants

All of the *Mtb* UgpB site-directed mutants were expressed and purified using the same optimised methods and steps as for UgpB-pYUB1062 (3.3.5) (3.3.7). SDS-PAGE and chromatography traces of all mutants can be found (Chapter 3, Appendix, Figures 3.9-3.17). All of the mutant derivatives of *Mtb* UgpB pYUB1062 (Table 3.1) were successfully expressed and purified however the yields of Asp102-Ala, Trp208-Ala and Gly306-Ala were markedly reduced compared to apo-UgpB. This could be due to the mutation altering the folding and stability of the protein. The fractions containing pure mutant derivatives of *Mtb* UgpB were combined, concentrated by ultrafiltration and stored in UgpB dialysis buffer 2 at -80 °C prior to analyses. Differences in the yields of mutant protein were due to protein precipitation during concentration (Table 3.1).

Table 3.1: Protein purification yields of *Mtb* UgpB mutants.

Mutant derivative of *Mtb* UgpB	Protein purification yield (mg per L)
Apo-UgpB	1.1
Tyr78Ala	2.6
Asp102Ala	0.6
Ser153Ala	2.1
Leu205Ala	1.2
Trp208Ala	0.2
Ser272Ala	2.0
Gly306Ala	0.3
Tyr345Ala	1.4
Arg385Ala	1.0

3.3.11 Circular dichroism analysis of *Mtb* UgpB mutants

Circular dichroism spectroscopy was used to analyse the secondary structure profile of all of the UgpB mutants. The proteins were dialysed into a low salt buffer (Chapter 2, Table 2.1). All of the measured mutants spectra showed a similar molar elipticity profile from 198-260 nm suggesting that the secondary structures were similar and that the substituted alanine mutation was not detrimental to the correct protein folding (Figure 3.8).

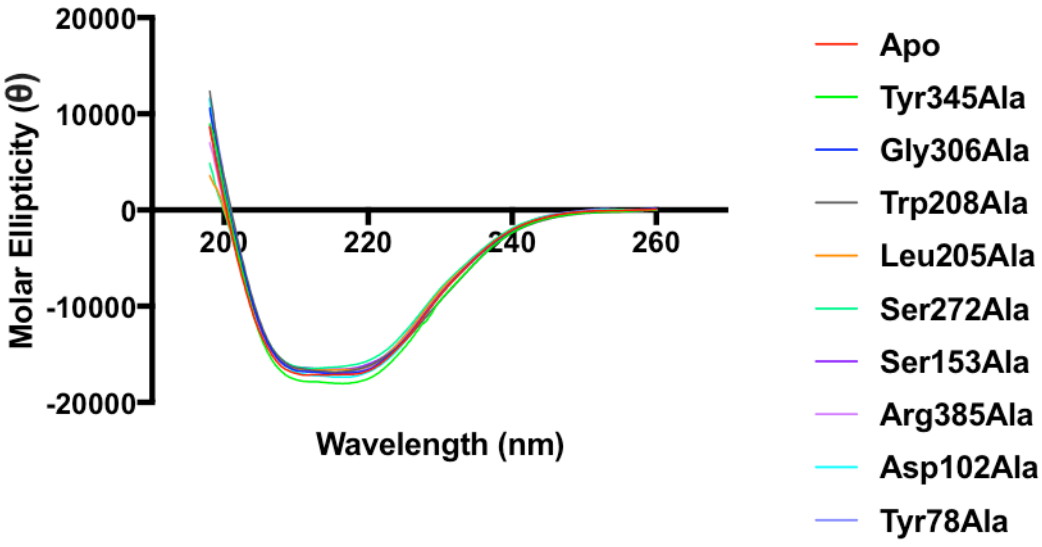


Figure 3.8: Circular dichroism spectra of *Mtb* UgpB and site directed mutant proteins.

3.4 Discussion

Mtb UgpB (Rv2833c) is the substrate binding protein of the *Mtb* UgpABCE transporter and was successfully purified to homogeneity with a yield of approximately 1 mg per L of culture. This was pivotal to carry out the biochemical and structural analyses as described (Chapter 4 and 5). In total 9 genetic variants of *Mtb* UgpB were also successfully expressed and purified with yields ranging from 0.2-2.6 mg per L of culture enabling further biochemical analyses (Chapter 4, 4.3.13). Lots of optimisations were previously attempted by members in the lab to express the *Mtb* UgpB protein including: different expression systems, vectors, host strains, full-length and truncated protein, homologues, different media and additives. However these were later found to not be successful and we found that *Mtb* UgpB needed to be expressed in *M. smegmatis* to produce soluble, non-aggregated protein. The levels of protein expression were comparable to other *Mtb* proteins expressed in *M. smegmatis* (Radhakrishnan et al., 2018) (Singhal et al., 2015). However we found that for the genetic variants the yield varied. This could be due to problems with protein folding and instability caused by substituting the native binding site residue for alanine.

Full length *Mtb* UgpB was expressed in *E. coli* in the pET28a vector under the control of the T7 promoter and containing a C-terminal hexa-histidine tag. Using this system *Mtb* UgpB could not be produced as soluble protein. This was probably because full length *Mtb* UgpB is predicted by TMpred (ExPASy) to form a single trans-membrane anchor helix and secreted signal sequence that may associate with the plasma membrane and cause it to be insoluble (Cuthbertson et al., 2005). In addition the potential lipobox motif may also contribute to the full-length protein being insoluble. Therefore to improve soluble expression of *Mtb* UgpB the first 34 amino acids at the N-terminus were removed, however soluble expression did not improve. In the previous study (Jiang et al., 2014) full-length *Mtb* UgpB was successfully expressed in *E.coli* BL21 (DE3) pLysS under control of the *tac* promoter with an N-terminal GST tag. The GST tag is reported to improve soluble protein expression (Rabhi-Essafi et al., 2007). Also *E. coli* pLysS cells express T7 lysozyme to prevent leaky expression from the T7 promoter (Rosano et al., 2014). We attempted to replicate this result however only found low levels of expression, the major product was free GST tag and most of the protein was insoluble. This data is not shown as a previous member in the lab group performed it. Therefore we decided to express both full length and truncated *Mtb* UgpB in pETSUMO together with the chaperone GroES (P. Gupta et al., 2006). The SUMO tag and GroES chaperone co-expression are reported to improve soluble expression and folding however, soluble expression was not obtained.

Different host organisms were used, including *Pseudomonas putida* (Ramos-Gonzalez et al., 2005) because it has a high G+C genome content similar to *Mtb* and may be particularly important for codon usage. Using the *Pseudomonas putida* host, only low levels of soluble *Mtb* UgpB expression were observed. Changing the host organism did appear to slightly improve soluble expression therefore it was instead expressed in the pYUB1062 vector and the more closely related *M. smegmatis* species under the control of the acetamidase promoter (Daugelat et al., 2003). In this system high levels of truncated *Mtb* UgpB expression were observed and soluble protein was successfully purified based on affinity of the C-terminal hexa-histidine tag for Ni^{2+} resin, followed by QHP ion exchange chromatography and Superdex 200 gel filtration chromatography. A TALON resin could be used to purify the full length UgpB protein without clarifying the cell lysate as physiologically it is a potential lipo protein.

Following the Ni^{2+} IMAC purification strategy, the protein was not completely pure and several low molecular weight contaminants remained. For biochemical and structural characterisation the protein needed to be as pure as possible therefore the purification steps had to be optimised. Instead of using Ni^{2+} resin, Co^{2+} resin was used. Co^{2+} generally has a lower affinity for histidine tags, however this can be advantageous as it then binds to the histidine tag and not histidine tag like proteins, which can result in a higher protein purity (ThermoFisher, 2019). Co^{2+} resin was found to result in purer protein before the first purification step. We also needed better resolution to separate *Mtb* UgpB from the lower molecular weight contaminants therefore instead of using a Superdex 200 gel filtration column a Superdex 75 gel filtration column was used. This was because the Superdex 75 gel filtration column has a lower molecular weight separation range of 3-70 kDa compared to 10-600 kDa. The combination of these optimisation steps resulted in the successful removal of most of the low molecular weight contaminants and purification of *Mtb* UgpB. We also mutated binding site residues of *Mtb* UgpB by site-directed mutagenesis. Following this optimised strategy, the *Mtb* UgpB site directed mutagenesis were successfully produced. Importantly the mutations didn't appear to alter protein solubility or stability based on circular dichroism analyses.

3.5 Appendix

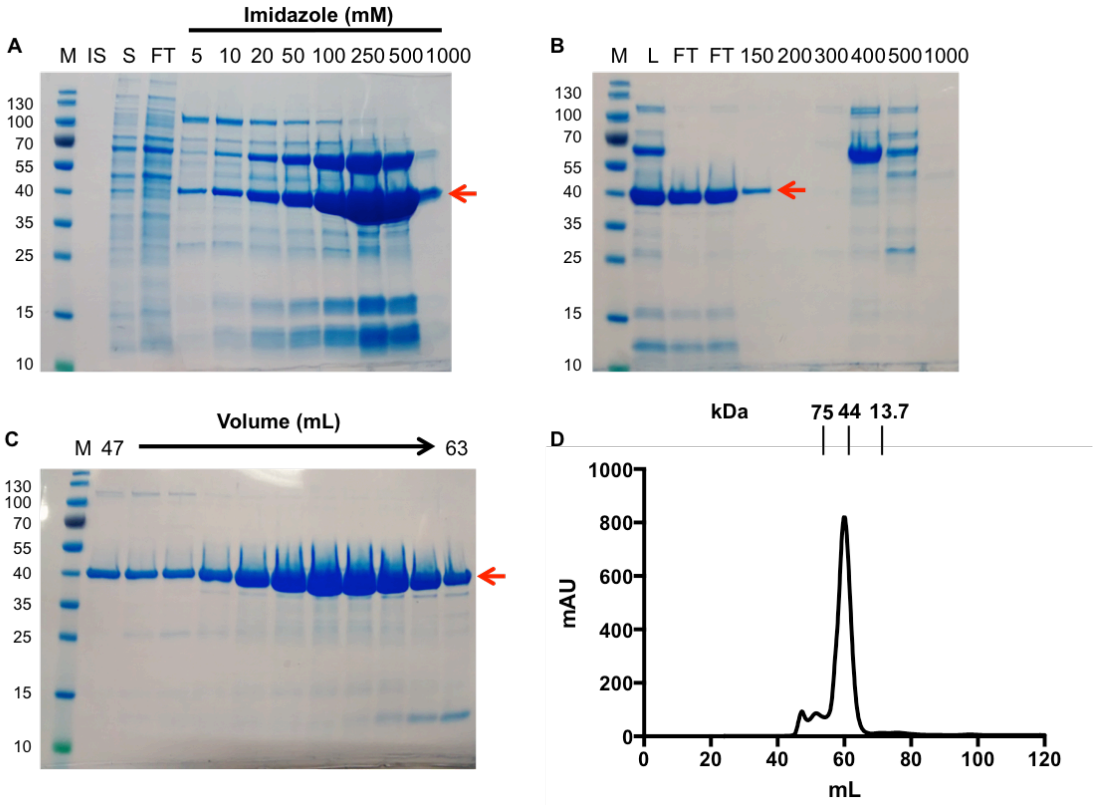


Figure 3.9: SDS-PAGE analysis of the purification of Tyr78Ala UgpB. A) Co^{2+} IMAC-column. B) QHP anion exchange chromatography. C) Superdex 75 gel filtration chromatography. D) Trace of C.

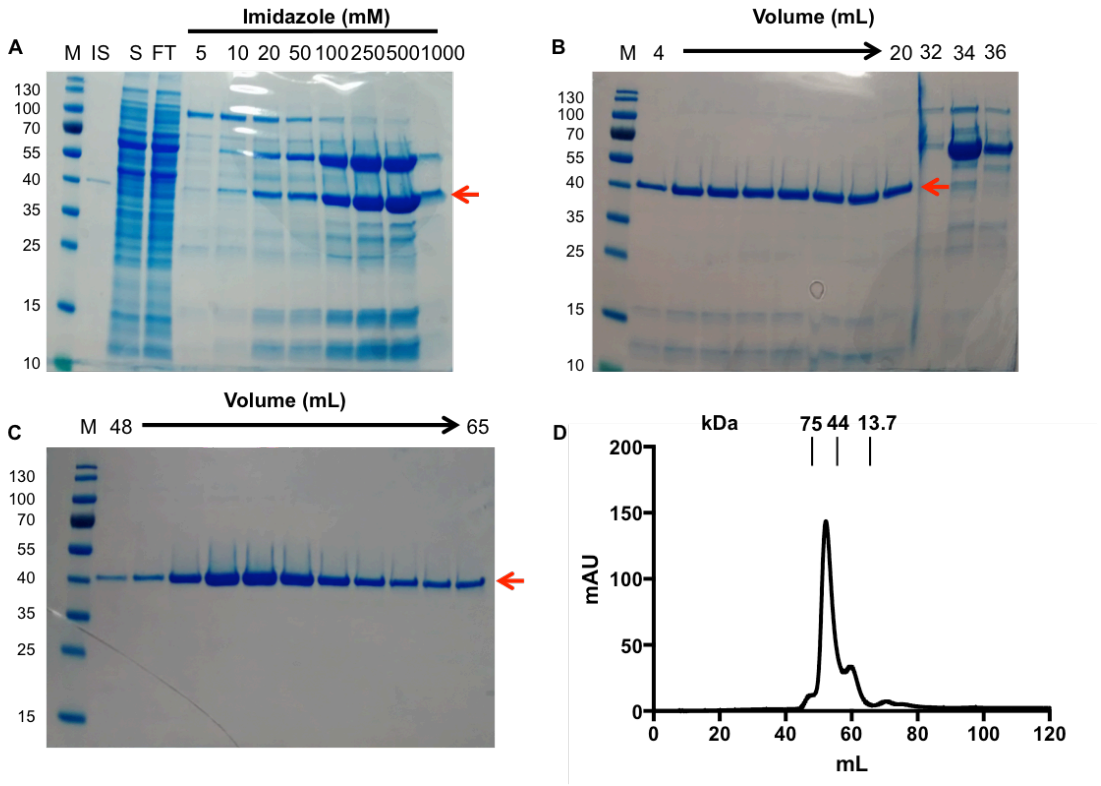


Figure 3.10: SDS-PAGE analysis of the purification of Asp102Ala UgpB. A) Co^{2+} IMAC-column. B) QHP anion exchange chromatography. C) Superdex 75 gel filtration chromatography. D) Trace of C.

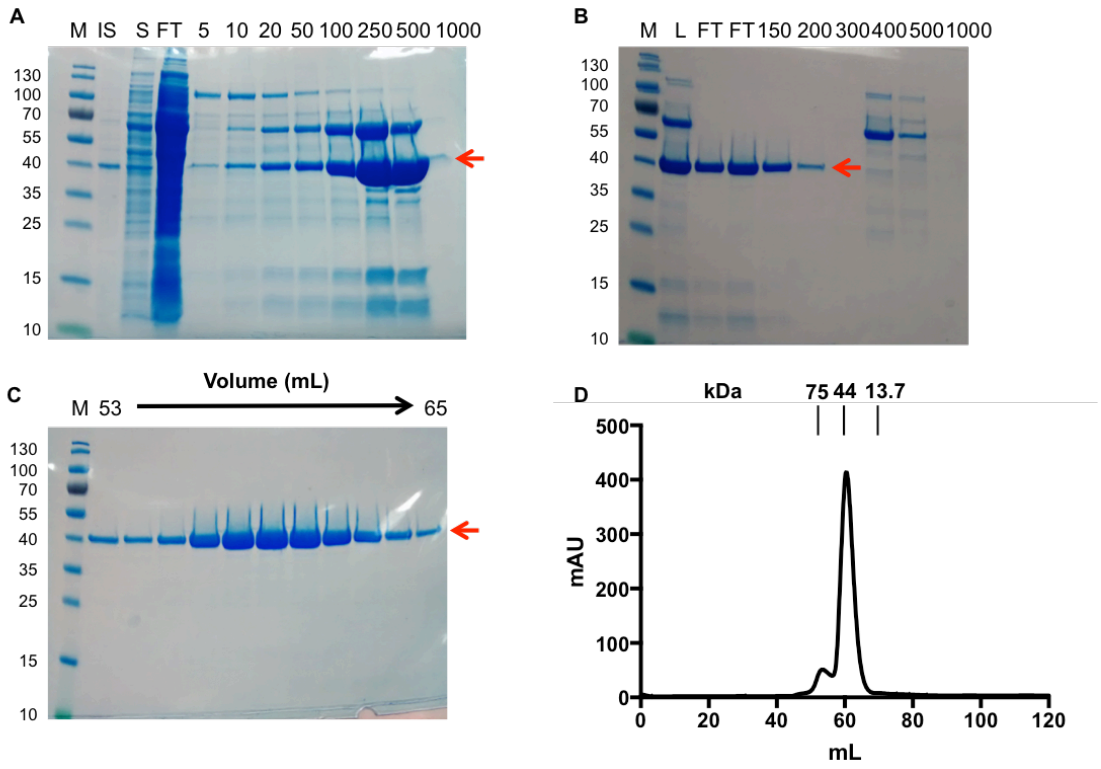


Figure 3.11: SDS-PAGE analysis of the purification of Ser153Ala UgpB. A) Co^{2+} IMAC-column. B) QHP anion exchange chromatography. C) Superdex 75 gel filtration chromatography. D) Trace of C.

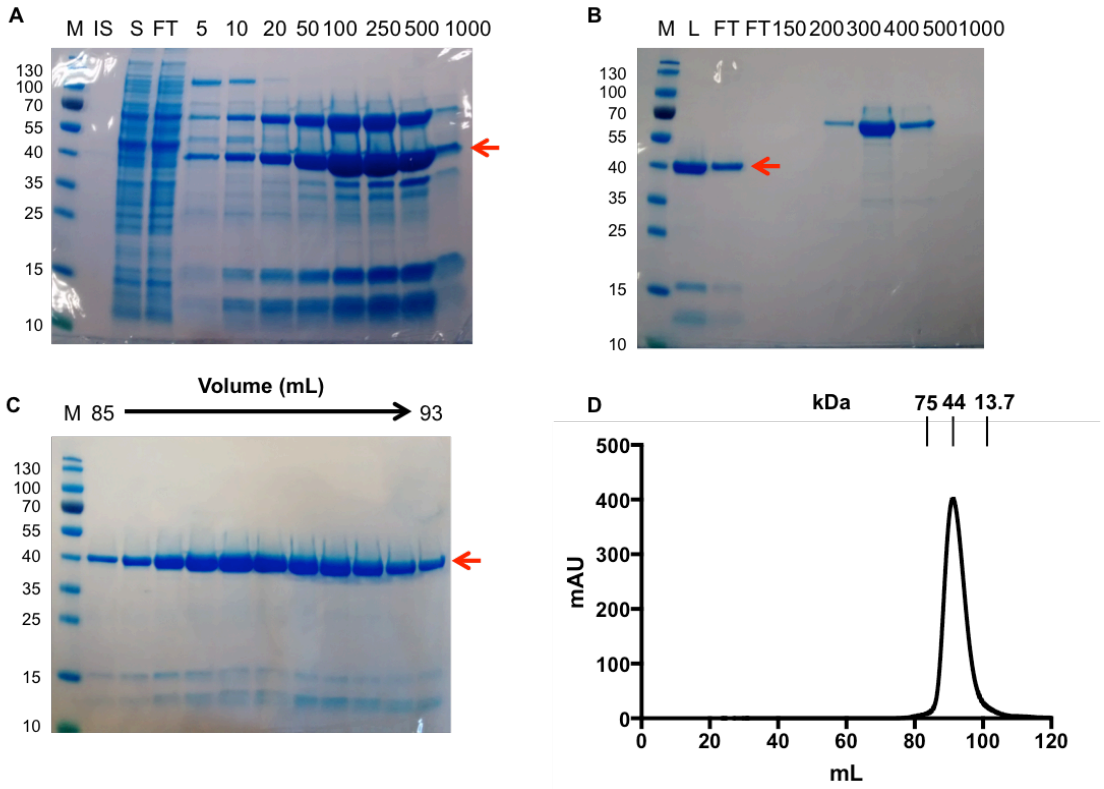


Figure 3.12: SDS-PAGE analysis of the purification of Leu205Ala UgpB. A) Co^{2+} IMAC-column. B) QHP anion exchange chromatography. C) Superdex 200 gel filtration chromatography. D) Trace of C.

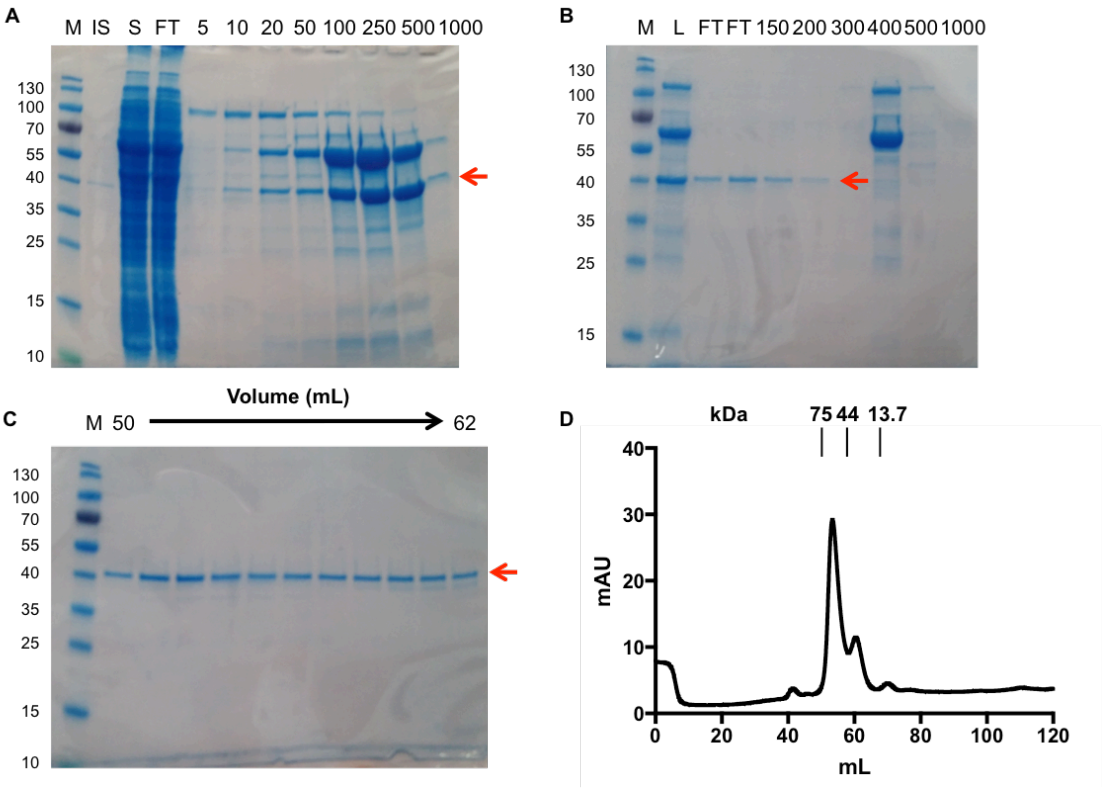


Figure 3.13: SDS-PAGE analysis of the purification of Trp208Ala UgpB. A) Co^{2+} IMAC-column. B) QHP anion exchange chromatography. C) Superdex 75 gel filtration chromatography. D) Trace of C.

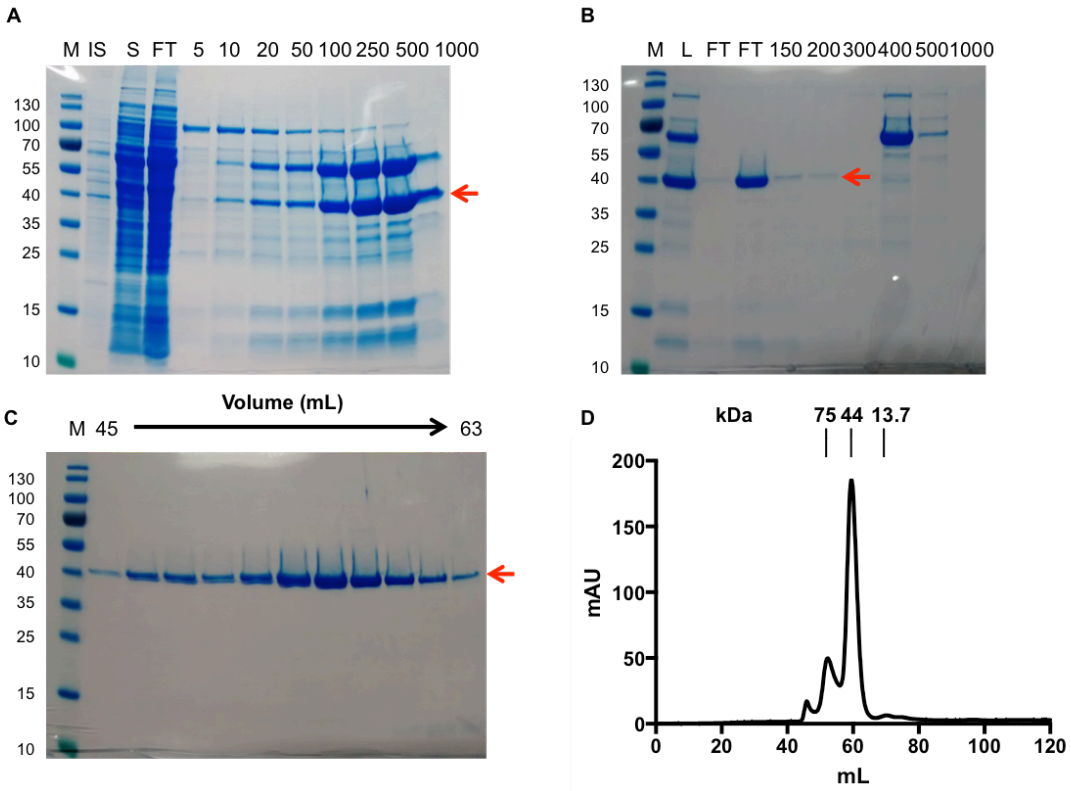


Figure 3.14: SDS-PAGE analysis of the purification of Ser272Ala UgpB. A) Co^{2+} IMAC-column. B) QHP anion exchange chromatography. C) Superdex 75 gel filtration chromatography. D) Trace of C.

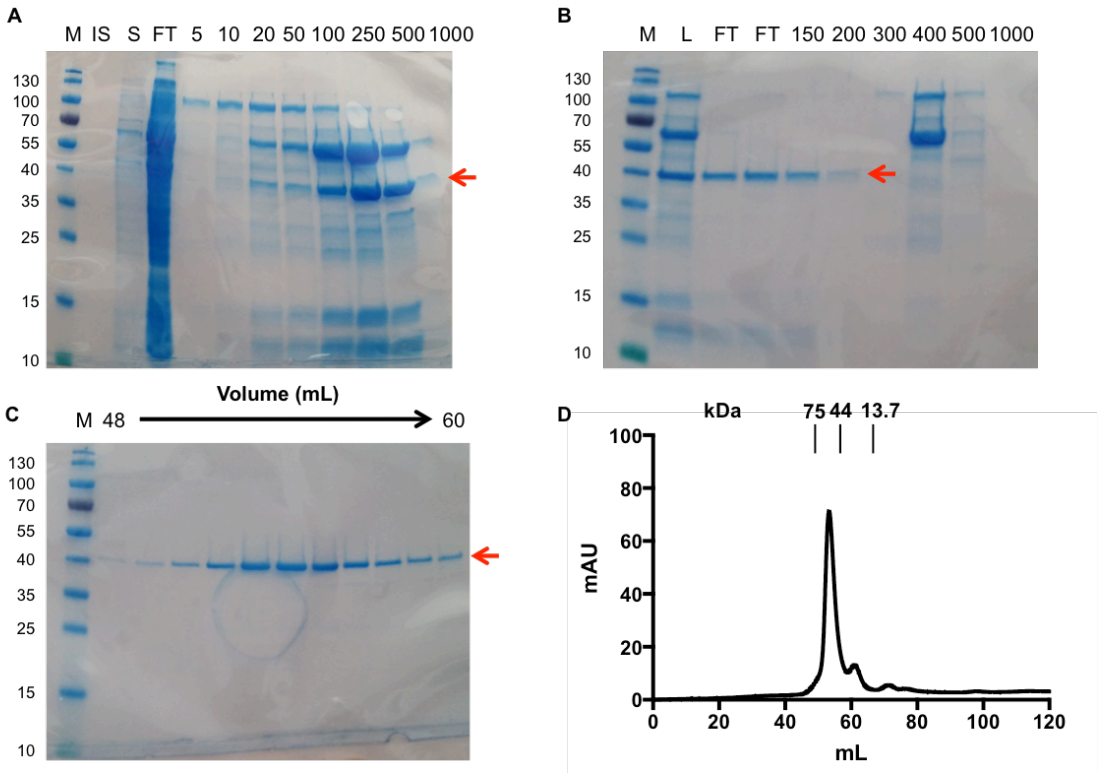


Figure 3.15: SDS-PAGE analysis of the purification of Gly306Ala UgpB. A) Co^{2+} IMAC-column. B) QHP anion exchange chromatography. C) Superdex 75 gel filtration chromatography. D) Trace of C.

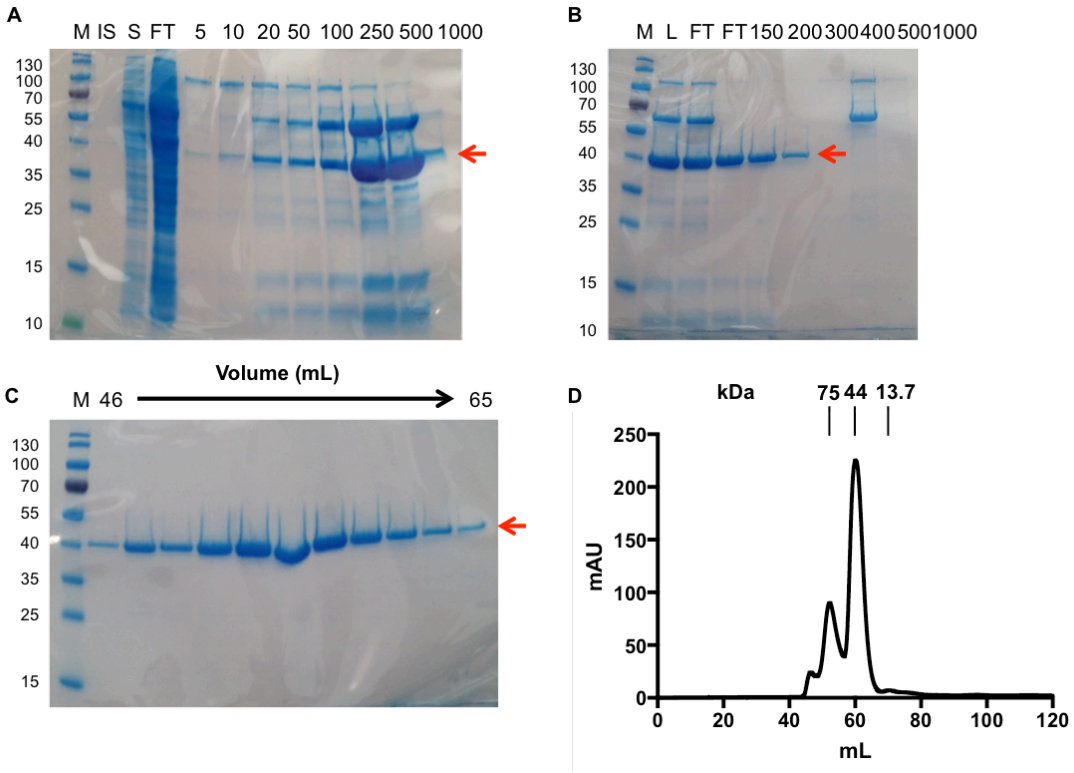


Figure 3.16: SDS-PAGE analysis of the purification of Tyr345Ala UgpB. A) Co^{2+} IMAC-column. B) QHP anion exchange chromatography. C) Superdex 75 gel filtration chromatography. D) Trace of C.

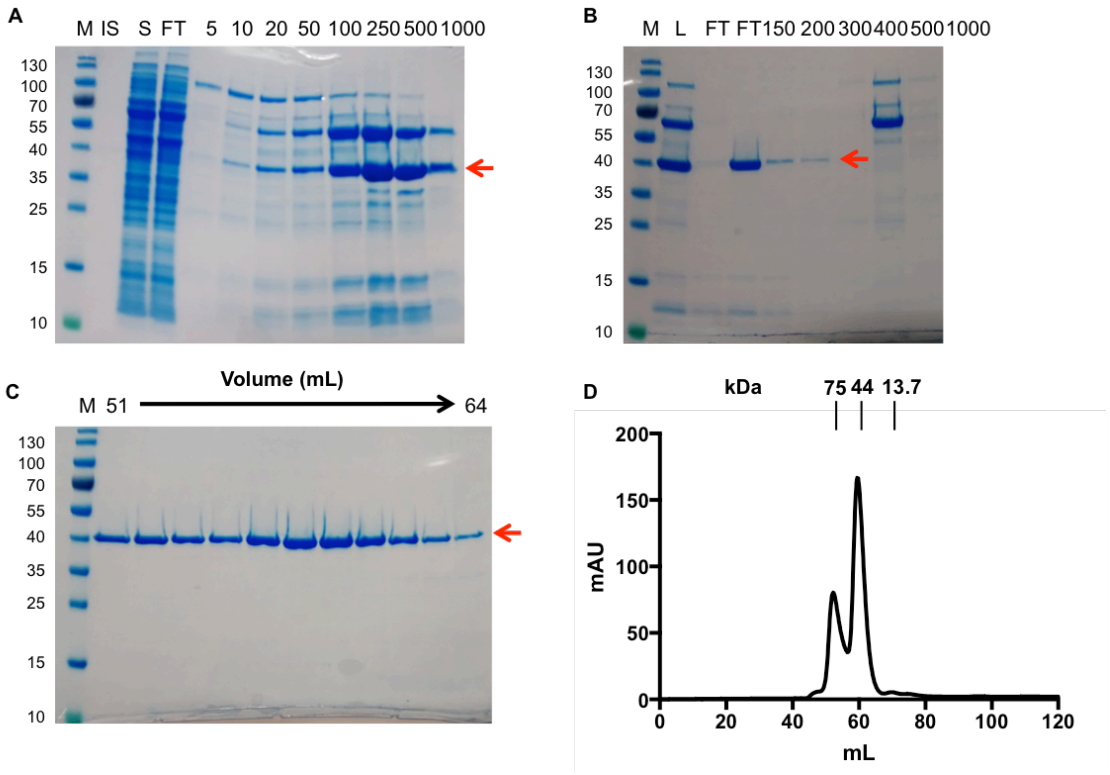


Figure 3.17: SDS-PAGE analysis of the purification of Arg385Ala UgpB. A) Co^{2+} IMAC-column. B) QHP anion exchange chromatography. C) Superdex 75 gel filtration chromatography. D) Trace of C.

Chapter 4: Characterisation of *Mycobacterium tuberculosis* UgpB

4.1 Introduction

4.1.1 Introduction

Mtb UgpB is the substrate binding protein (SBP) that is part of the UgpABCE ATP-binding cassette (ABC) transporter. As *Mtb* UgpB was successfully purified to homogeneity in Chapter 3 the next step was to biochemically characterise its ligand interactions. Only a limited number of molecules have been investigated for binding to *Mtb* UgpB (Jiang et al., 2014) therefore we set out to carry out a more extensive biochemical analysis.

4.1.2 Ligand binding studies of *Mtb* substrate binding proteins

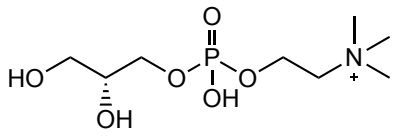
The typical binding affinities of SBP's range from nM to mM range (Berntsson et al., 2010). As ligand binding to SBP's can be difficult to measure, several different methods can be used. Methods to investigate SBP ligand binding include: radioactive assays (Pandey et al., 2016), surface plasmon resonance (SPR) (Kumar et al., 2019), intrinsic tryptophan fluorescence (Bhattacharyya et al., 2018), isothermal calorimetry (Wuttge et al., 2012), Forster resonance energy transfer (FRET) (Deuschle et al., 2005) and microscale thermophoresis (MST) (Brautigam et al., 2014). However only a limited number of studies have biochemically characterised *Mtb* SBP's.

Recently, the putative *Mtb* glutamine SBP, GlnH was biochemically investigated (Bhattacharyya et al., 2018). Despite being annotated as a glutamine binding protein *Mtb* GlnH was actually found to have the highest binding affinity to aspartate (K_d = 4.8 μM compared to 15.2 μM) by isothermal calorimetry (ITC). In contrast, a GlnH homologue sharing only 43 % amino acid sequence similarity from *Corynebacterium glutamicum* had a significantly lower affinity to aspartate (K_d = 550 μM) and glutamate (K_d = 2,060 μM) by intrinsic tryptophan fluorescence (Bhattacharyya et al., 2018). Clearly this study demonstrates the requirement to functionally investigate SBP's that are predicted by bioinformatics to have similar functions. Another example is the putative compatible solute-binding protein, *Mtb* ProX, which was predicted by bioinformatics to bind to the osmolytes: glycine and betaine however a functional study of *Mtb* ProX found that it actually binds to polyphenols and may be involved in the uptake of polyphenol analogues during infection (J. H. Zhao et al., 2018).

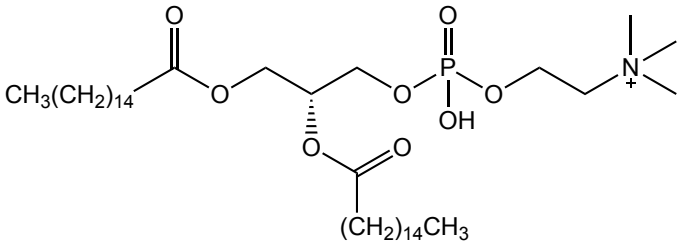
SBP's that bind carbohydrates typically have lower binding affinities compared to amino acid or metal ion specific SBP's (Phillips et al., 1976) (Cuneo et al., 2006) (Yatsunyk et al., 2008) (Pulido et al., 2015). The *Mtb* LpqY-SugABC transporter has been found to be involved in uptake of the disaccharide trehalose, however the binding affinity of trehalose to the SBP *Mtb* LpqY has not yet been determined, and other ligands may exist (Kalscheuer et al., 2010). A predicted carbohydrate SBP, *Mtb* UspC was recently investigated for its ability to bind a panel of carbohydrates by a thermal shift assay and found to preferentially bind to amino-sugars: D-glucosamine-6-phosphate (K_d = 27.7 mM) and chitobiose (K_d = 38.1 mM) (Fullam et al., 2016). Complementary NMR analysis revealed the structural basis of these carbohydrate interactions and highlighted a potential role of *Mtb* UspC in the uptake of peptidoglycan cell wall fragments (Fullam et al., 2016). Similarly this study highlights the importance of screening a wide range of potential substrates to investigate SBP function. Importantly, biochemical studies of *Mtb* SBP's may reveal different nutrients and regulatory pathways utilised by *Mtb* during infection and could lead to the identification of novel targets for drug development.

4.1.3 Ligand binding studies of *Mtb* UgpB and *E. coli* UgpB

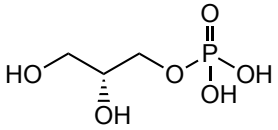
A previous study of *Mtb* UgpB (Jiang et al., 2014) identified binding to glycerol-3-phosphocholine (GPC) (Figure 4.1), the head-group of the phospholipid phosphatidylcholine (Figure 4.1) by ITC. However the study (Jiang et al., 2014) was limited in scope as they only tested binding to four molecules: GPC, glycerol-3-phosphate (G3P), glycerol-2-phosphate (G2P) and maltose, the structures are shown (Figure 4.1). The binding affinity to glycerol-3-phosphocholine (1:1 cadmium chloride adduct) was determined with a K_d of 27.3 µM however binding to the linear structurally related G3P, the branched G2P and structurally dissimilar maltose could not be detected. Despite the high affinity to GPC the authors were unable to co-crystallise *Mtb* UgpB with GPC (Jiang et al., 2014). In comparison, *E. coli* UgpB has a low amino acid sequence similarity (26.4 % similarity over 83 % read) (Chapter 3, Figure 3.2) compared to *Mtb* UgpB however for *E. coli* UgpB, binding was detected to both G3P with a K_d of 0.68 µM and also to GPC with a K_d of 5.1 µM by ITC (Wuttge et al., 2012). The observation of binding of GPC to both *Mtb* UgpB and *E. coli* UgpB was puzzling as both of these proteins do not have a choline-binding motif such as that found in the choline/acetylcholine binding protein aromatic box of *S. meliloti* ChoX (Oswald et al., 2008). We therefore wanted to investigate if other potential ligands bind to *Mtb* UgpB and also determine the molecular basis of substrate recognition (Chapter 5).



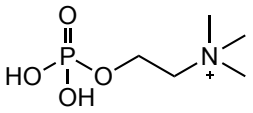
Glycerol-3-phosphocholine



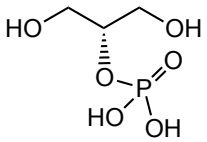
Phosphatidylcholine



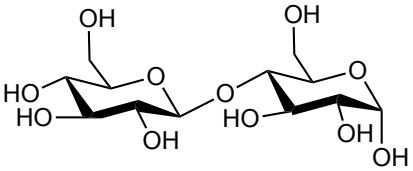
Glycerol-3-phosphate



Phosphocholine



Glycerol-2-phosphate



maltose

Figure 4.1: Structures of glycerophosphodiester variations investigated for binding to *Mtb* UgpB and *E. coli* UgpB (Jiang et al., 2014) (Wuttge et al., 2012)**.**

4.2 Hypotheses and Aims

I hypothesised that the *Mtb* UgpB SBP binds to alternative ligands. The protein is predicted by comparative genomics to be a carbohydrate binding protein, and the aim here was to test binding to a range of different substrates. Many ABC transporters are promiscuous therefore I also tested binding to amino acids, anti-tubercular antibiotics and different components of glycerophosphodiesters. In order to test binding I used a thermal shift assay. A previous study found that the protein binds to a glycerophosphodiester called GPC. However I hypothesised that alternative glycerophosphodiesters that possess a glycerol tail and phosphate group, potentially derived from macrophage membranes could also bind. Finally, I hypothesised that site-directed mtuants would have different glycerophosphodiester binding affinities compared to the apo protein. To test these hypotheses I synthesised alternative glycerophosphodiesters and tested binding by MST.

The specific aims of the work presented in this chapter were:

- To carry out a ligand binding screen to biochemically characterise *Mtb* UgpB.
- To establish if alternative lipid head groups are recognised by *Mtb* UgpB.
- To determine the binding affinities of *Mtb* UgpB for the ligands identified.
- To determine the effect of site directed mutants on binding and ligand recognition.

4.3 Results

4.3.1 Introduction

Mtb UgpB is known to bind to GPC (Jiang et al., 2014) To explore whether *Mtb* UgpB recognises other ligands *Mtb* UgpB was expressed as pure protein (Chapter 3, 3.3.7) and the biochemical characterisation is described in this chapter. The ligands identified for *Mtb* UgpB in this chapter have subsequently been used to set crystal trials to investigate the structural determinants of these interactions (Chapter 5).

4.3.2 Thermal shift assay of *Mtb* UgpB

To investigate if other ligands bind to *Mtb* UgpB a thermal shift assay was carried out as described (Chapter 2, 2.5.8). The thermal shift assay works by measuring the excitation/emission wavelength profile of SYPRO Orange fluorescent dye (Niesen et al., 2007). The T_m is calculated by solving the first derivative of the fluorescence emission as a function of temperature and identifying the lowest point of the curve (Huynh et al., 2015). Comparing the T_m of a protein in the presence or absence of ligand is used to calculate the ΔT_m. In this study, a panel of carbohydrates, amino acids, and anti-mycobacterial antibiotics were selected because they were readily available, some were known to be ligands based on previous studies (Jiang et al., 2014) and carbohydrates were tested because of the bioinformatics sequence similarity of *Mtb* UgpB to other carbohydrate binding proteins (Niederweis, 2008).

All of the carbohydrates were tested at a final concentration of 100 mM however the amino acids were tested at the highest soluble concentration. The anti-mycobacterial antibiotics were tested at a final concentration of 10 mM. In each assay, *Mtb* UgpB was tested at a final concentration of 22 µM in the buffer (150 mM NaCl, 25 mM HEPES, 10 % glycerol pH 7.0). 15 X final concentration of SYPRO orange dye was added together with the ligand in a total volume of 20 µL and incubated at room temperature for 15 min prior to measurement.

4.3.3 Thermal shift assay with *Mtb* UgpB and carbohydrates

A total of 23 carbohydrate compounds were tested and the structures that were found to bind are shown (Figure 4.2). A range of different carbohydrates was tested including monosaccharides, disaccharides, amino-sugars and carbohydrates that resembled peptidoglycan cell wall components. Trehalose was tested because it is known to bind to a different *Mtb* substrate binding protein: LpqY that belongs to the LpqY-SugABC transporter (Kalscheuer et al., 2010). Amino-sugars were tested because one of the

other *Mtb* predicted carbohydrate transporters UspC was found to bind to D-glucosamine-6-phosphate and chitobiose (Fullam et al., 2016).

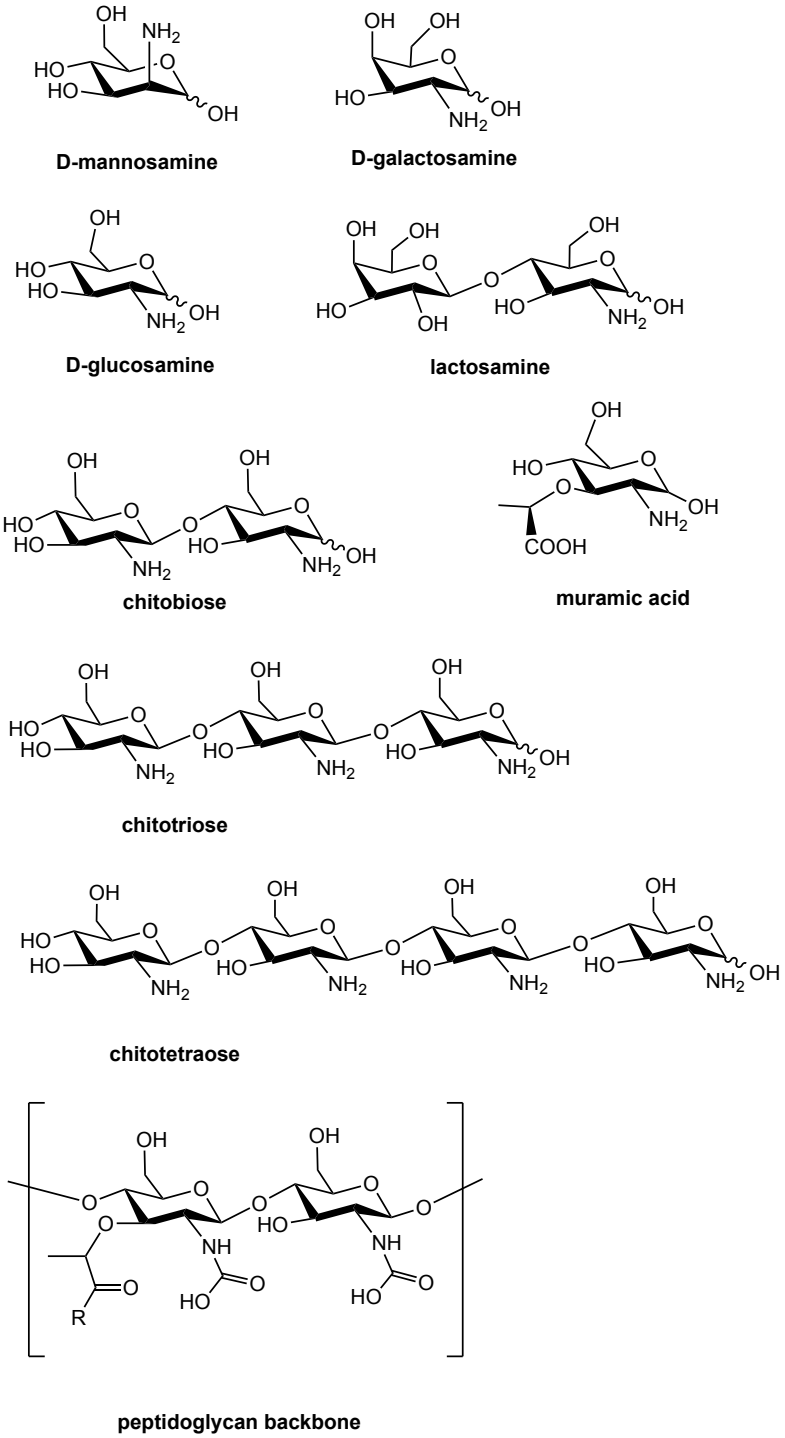


Figure 4.2: Structures of carbohydrates that indicated binding to *Mtb* UgpB by thermal shift assay.

The results of the thermal shift assay with *Mtb* UgpB and carbohydrates (Figure 4.3) indicated that 8 amino sugars appeared to produce the highest thermal shifts: D-mannosamine (+4.8 °C), D-galactosamine (+4.6 °C), D-glucosamine (+4.4 °C) and lactosamine (+3.6 °C). The disaccharide amino sugar, lactosamine showed a similar

ΔT_m to the monosaccharide amino sugars therefore size of the sugar didn't appear to be a determinant of binding. Other carbohydrates that produced only small positive thermal shifts were the amino sugars, muramic acid (+2.6 °C), chitobiose (+2.8 °C), chitotriose (+2.4 °C) and chitotetraose (-2.15 °C) which are all carbohydrates that resemble peptidoglycan cell wall components (Figure 4.2). Ligand binding to the most populated, lowest energy, native protein state will cause stabilisation and results in a positive thermal shift value. However, the protein is conformationally dynamic and exists in different energy states. If the ligand binds to a less populated, conformational state, ligand binding may result in a negative thermal shift value (Cimmperman et al., 2008). Binding to amino-sugars was similar to a previous study of *Mtb* UspC that showed binding to D-glucosamine-6-phosphate (+3.0 °C) however the thermal shift for the disaccharide chitobiose was a lot higher for *Mtb* UspC (+6.7 °C) (Fullam et al., 2016).

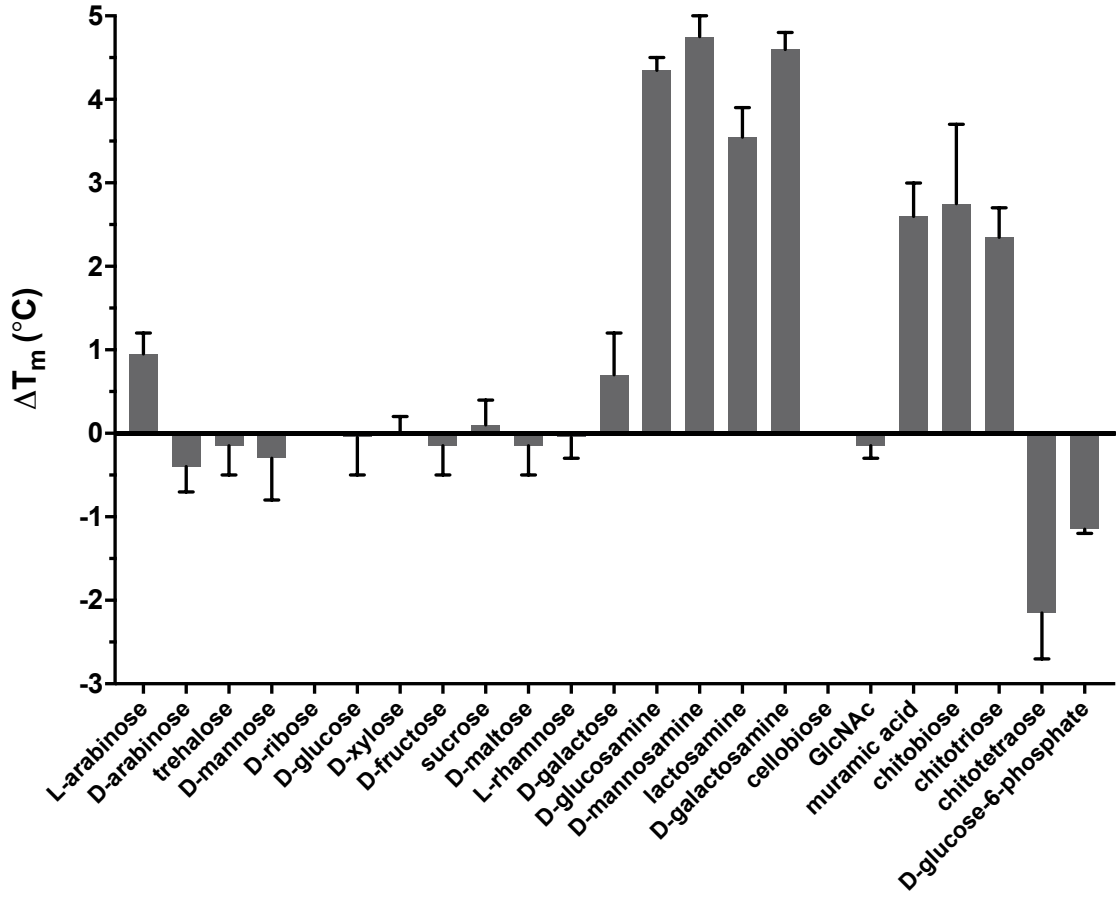


Figure 4.3: Thermal shift assay with *Mtb* UgpB and carbohydrates. Bar graphs illustrating thermal shifts (ΔT_m) defined as the difference in thermal shift of protein with ligand compared to without ligand. The different ligands were probed for binding at a final concentration of 100mM. Data are shown from three independent repeats represented as mean ± SD.

4.3.4 Thermal shift assay with *Mtb* UgpB and variations of glycerol-3-phosphocholine

Different variations of GPC were also tested because a previous study showed that *Mtb* UgpB binds to GPC:$CdCl_2$ (1:1 cadmium chloride adduct) (Jiang et al., 2014). Structures of all the molecules tested are shown (Figure 4.4). A previous study of *E. coli* UgpB also identified that it binds to both G3P and GPC (Wuttge et al., 2012) so we wanted to see if this could be observed for *Mtb* UgpB using the thermal shift assay. Therefore GPC and G3P (Figure 4.4) were tested for binding as well as other variations of GPC.

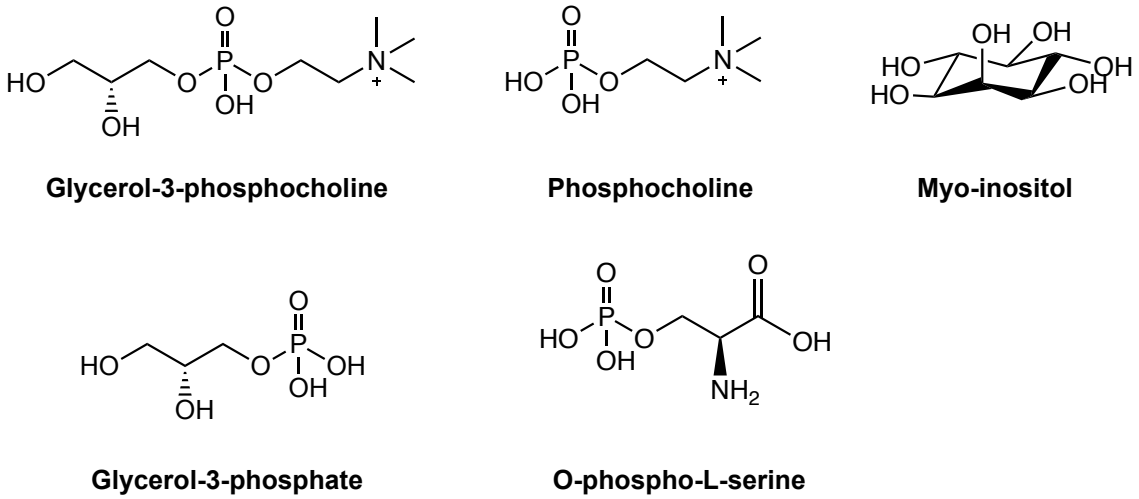


Figure 4.4: Structures of glycerophosphodiesters and analogues probed by thermal shift assay.

The results of the thermal shift assay with *Mtb* UgpB and variations of GPC (Figure 4.5) indicated that GPC produced the highest thermal shift (+7.3 °C) which supports the previous study that revealed GPC bound with a high affinity K_d = 27.3 μM by isothermal calorimetry (Jiang et al., 2014). However, in our studies the GPC:$CdCl_2$ (1:1 cadmium chloride adduct) indicated a much lower thermal shift (-11.7 °C). The previous study (Jiang et al., 2014) also tested GPC:$CdCl_2$ (1:1 cadmium chloride adduct) binding to *Mtb* UgpB and our thermal shift results suggest it may actually have a destabilising effect. This is further supported as when $CdCl_2$ was tested it produced a large negative thermal shift (-28.0 °C). Therefore GPC that is available as a non-$CdCl_2$ adduct was tested, and it was found to produce the highest thermal shift (+7.3 °C). This suggests that GPC is binding, and $CdCl_2$ is not binding. Other derivatives of GPC include phosphocholine (Figure 4.4) that lacks the glycerol moiety of GPC and G3P (Figure 4.4) that lacks the phosphocholine moiety of GPC therefore we wanted to test these compounds to investigate what GPC moieties were required for binding.

Phosphocholine (Figure 4.5) only produced a small negative thermal shift (-1.4 °C) suggesting that the glycerol moiety of GPC is important for binding. G3P (Figure 4.4) was also tested because it showed binding to *E. coli* UgpB with a K_d of 0.68 μM by isothermal calorimetry (Wuttge et al., 2012). G3P has a glycerol tail and lacks a choline head-group (Figure 4.4). However G3P only showed a small negative thermal shift (-1.4 °C) that contrasted with the previous study of *E. coli* UgpB however agreed with the *Mtb* UgpB study by isothermal calorimetry (ITC) (Jiang et al., 2014). We also wanted to investigate whether head group type was important. O-phospho-L-serine (Figure 4.4), the head-group of phosphatidyl-L-serine without a glycerol backbone only showed a small negative thermal shift of (-3.3 °C) whereas myo-inositol the head-group of phosphatidylinositol showed a very small negative thermal shift (-0.2 °C) however myo-inositol does not contain a glycerol or phosphate group. Later we show that these head groups are recognised by *Mtb* UgpB (Chapter 4, 4.5) and that they require a glycerol moiety for binding recognition. Combined these results suggested that the intact glycerol, phosphate and choline moiety together in GPC were required for binding.

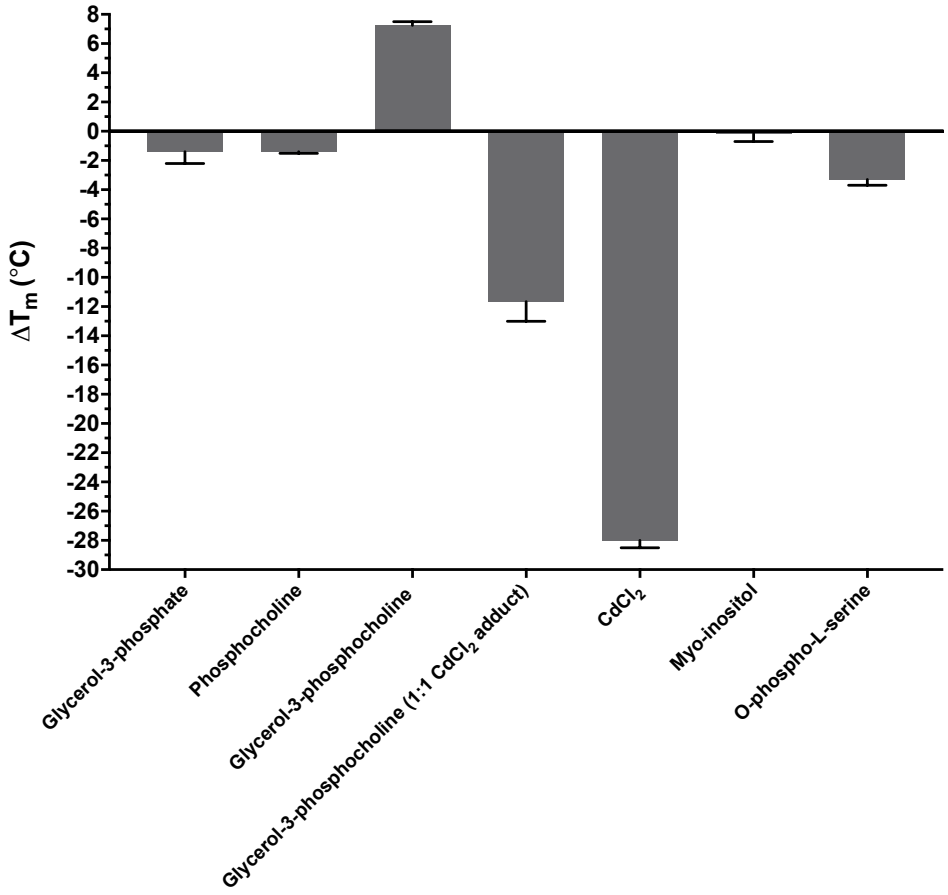


Figure 4.5: Thermal shift assay with *Mtb* UgpB and glycerophosphodiester variations. Bar graphs illustrating thermal shifts of ΔT_m defined as the difference in thermal shift of protein with ligand compared to without ligand. The different ligands were probed for binding at a final concentration of 100mM. Data are shown from three independent repeats represented as mean ± SD.

4.3.5 Thermal shift assay with *Mtb* UgpB and amino acids

To evaluate if *Mtb* UgpB binds to amino acids the thermal shift assay was carried out using the same conditions as described (4.3.2) (Chapter 2, 2.5.8). The amino acids were tested at the highest soluble concentration of 100 mM apart from those indicated (Figure 4.6) due to their low solubility in water.

The results of the amino acid analysis (Figure 4.6) indicated that arginine produced the largest negative thermal shift (-3.8 °C). The structures of amino acids that indicated binding are shown (Figure 4.7). Arginine is frequently added to stabilise proteins from aggregation in refolding experiments and may explain this result as the thermal shift assay measures protein unfolding (Baynes et al., 2005). Histidine produced the largest positive thermal shift (+3.7 °C) and as well as arginine it has a positively charged side chain. As *Mtb* UgpB has a C-terminal hexa-histidine tag that was not removed following purification the histidine may be interacting with the intrinsic hexa-histidine tag, altering its stability and hence the observed thermal shift. However there is no cleavage site in

Mtb UgpB to remove the C-terminal hexa-histidine tag therefore this hypothesis could not be tested. Two negatively charged amino acids, glutamate and aspartate produced the second largest thermal shifts, glutamate (+3.0 °C) and aspartate (+1.8 °C) whereas the other classes of amino acids did not produce any ΔT_m > 2.0 °C therefore it appears that charged amino acids in particular affected the thermal shift.

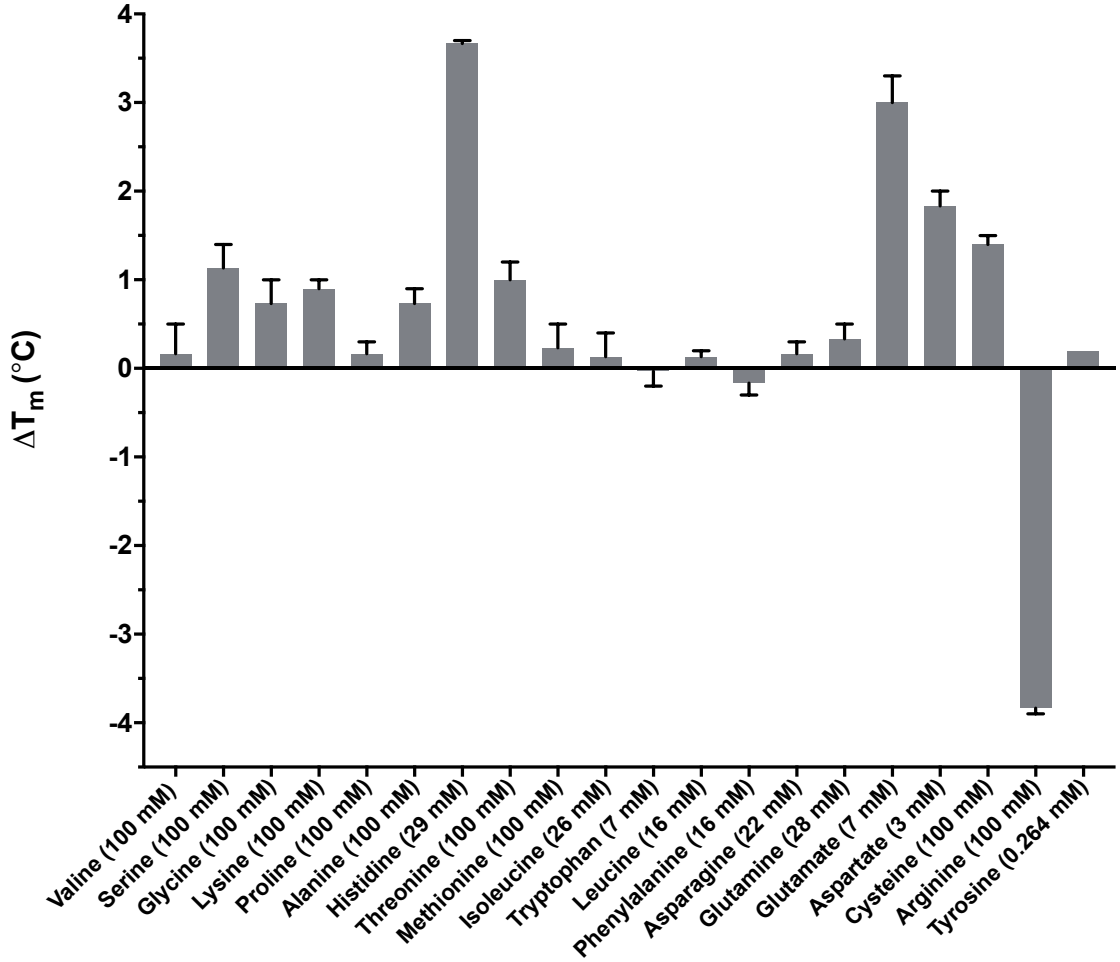


Figure 4.6: Thermal shift assay with *Mtb* UgpB and amino acids. Bar graphs illustrating thermal shifts of ΔT_m defined as the difference in thermal shift of protein with ligand compared to without ligand. The different compounds were probed for binding at the final concentrations in brackets. Data are shown from three independent repeats represented as mean ± SD.

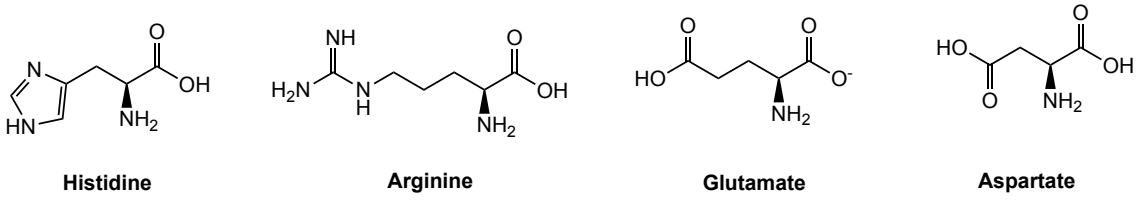


Figure 4.7: Structures of amino acids that indicated binding to *Mtb* UgpB by thermal shift assay.

4.3.6 Thermal shift assay with *Mtb* UgpB and antibiotics

In order for antibiotics to be transported into cells they must either cross the lipid bilayer of the cell membrane directly or pass through protein transporters. However recent studies have indicated that there is little evidence of drug transport through the phospholipid bilayer and instead protein transporters mediate drug transport (Kell, 2015) (Kell et al., 2013). Therefore we decided to test if a range of anti-mycobacterial antibiotics including some of those routinely used to treat tuberculosis could bind to *Mtb* UgpB by the thermal shift assay.

The antibiotics were tested at a final concentration of 10 mM and the thermal shifts are shown (Figure 4.8). Interestingly three of the six antimycobacterial antibiotics produced large thermal shifts (Figure 4.8). Ethambutol produced the highest positive thermal shift (+3.7 °C). The structure of ethambutol resembles GPC (Figure 4.9) however ethambutol only has two hydroxyl groups instead of three hydroxyl groups in GPC. The two hydroxyl groups in ethambutol may be important for binding because based on our GPC bound structure the three hydroxyl groups of GPC are important for binding recognition (Chapter 5, 5.3.6). Hygromycin produced the largest negative thermal shift (-4.2 °C) however isoniazid and pyrazinamide only produced very small thermal shifts. As the antibiotics were tested at a final concentration of 10 mM, GPC was also tested at 10 mM and showed only a small thermal shift (+1.0 °C) compared to the large thermal shift (+7.3 °C) at 100 mM. Interestingly these results may indicate that *Mtb* UgpB binds to antibiotics and could have a role in the uptake of these compounds. Further work in the lab will investigate this further.

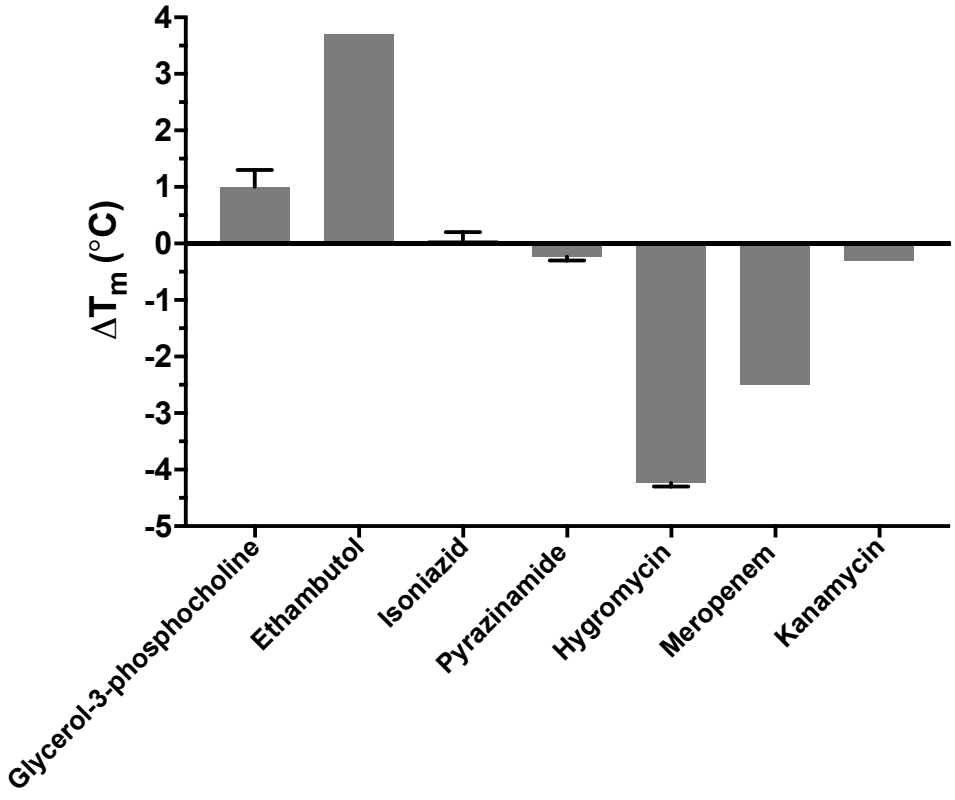
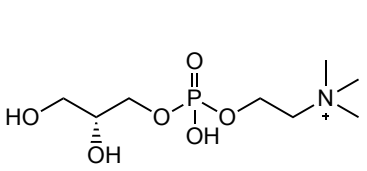
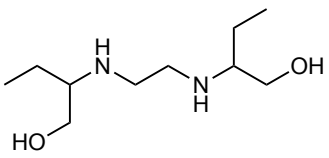


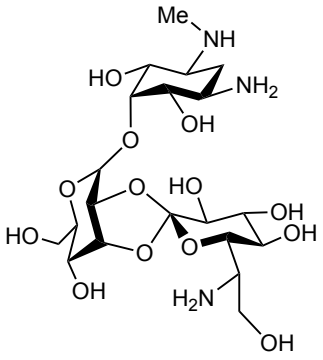
Figure 4.8: Thermal shift assay with *Mtb* UgpB and antibiotics. Bar graphs illustrating thermal shifts of ΔT_m defined as the difference in thermal shift of protein with ligand compared to without ligand. The different compounds were probed for binding at final concentrations of 10 mM. Data are shown from three independent repeats represented as mean ± SD.



Glycerol-3-phosphocholine



Ethambutol



Hygromycin

Figure 4.9: Structures of ethambutol and hygromycin that indicated binding to *Mtb* UgpB by thermal shift assay.

4.3.7 *Mtb* UgpB phospholipid strips

Based on the thermal shift results that indicated *Mtb* UgpB binds to GPC (Figure 4.5) we wanted to evaluate if other lipid head-groups could also bind to *Mtb* UgpB. A convenient method to do this is to measure binding to the phospholipids e.g. phosphatidylcholine is the lipid with a choline head group (Figure 4.10). Phospholipid

strips are commercially available (Echelon Biosciences) and are nitrocellulose membranes spotted with 100 pmol of lipid per spot.

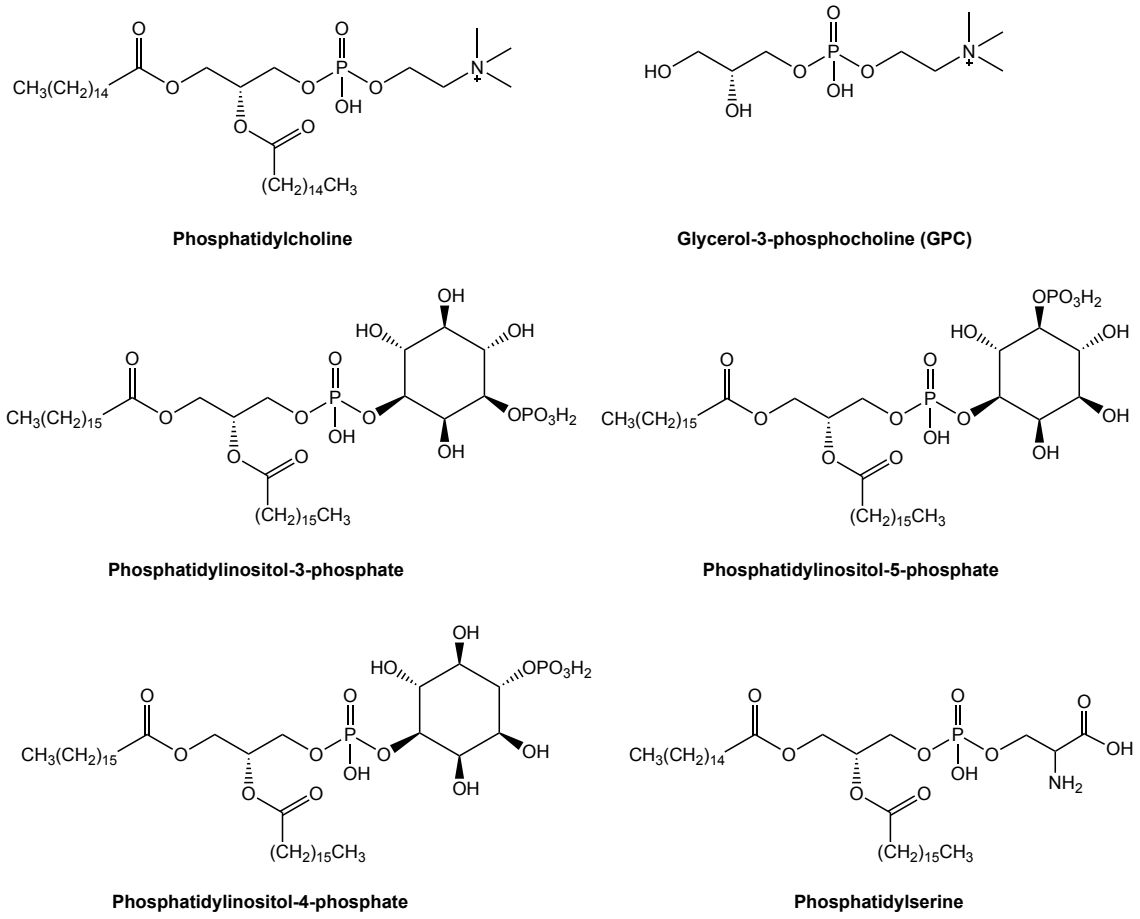


Figure 4.10: Structures of phospholipids that indicated the strongest binding to *Mtb* UgpB by phospholipid strip. The structure of glycerol-3-phosphocholine (GPC) is shown as it is derived from phosphatidylcholine.

To determine the binding of *Mtb* UgpB to lipids the strips were treated as described in the methods (Chapter 2, 2.5.9). *Mtb* UgpB was incubated with the lipid strip and binding to the lipid spots was detected by incubating with penta-His primary antibody, followed by incubating with anti-mouse IgG-HRP secondary antibody. Chemiluminescence was used to detect the bound antibody. The membranes were quantified by calculating the integrated density of spots in ImageJ (Fiji).

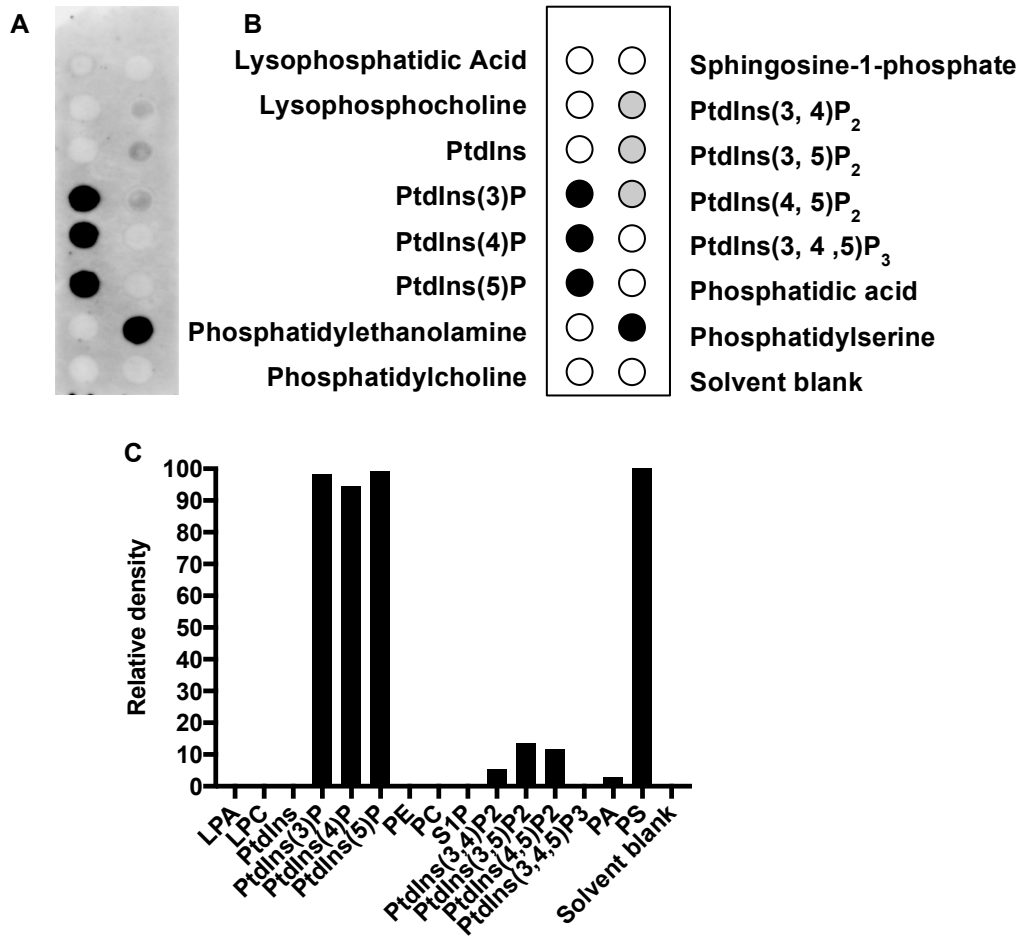


Figure 4.11: Phospholipid strip *Mtb* UgpB. A) Phospholipid strip probed with *Mtb* UgpB and anti-His/anti HRP secondary and ECL reagent. B) Schematic of phospholipid strip, black indicates strongest binding, grey weaker binding and white no binding. C) Quantitation of phospholipid strip binding calculated in ImageJ (n=1).

The results of the phospholipid strip (Figure 4.11) indicated strongest binding of *Mtb* UgpB to inositol phosphate lipids containing a single phosphate group: Phosphatidyl-inositol-3-phosphate (PtdIns(3)P), phosphatidyl-inositol-4-phosphate (PtdIns(4)P) and phosphatidylinositol-5-phosphate (PtdIns(5)P) (Figure 4.11). The results also indicate that *Mtb* UgpB binds to phosphatidylserine (PS). These results were intriguing as it suggested that potentially alternative phospholipid head-groups could also bind to *Mtb* UgpB. However no binding was observed to phosphatidylcholine (PC) despite it having the GPC head-group. These phospholipid strip assays are qualitative and were only carried out once. The assays would need to be repeated at least 3 times to show significance of the data. However these assays are renowned for false positives and negatives (Downes et al., 2005). It may be that as the glycerol moiety is ester bonded to two fatty acid tails in PC (Figure 4.10) this may inhibit binding recognition and based on the structure of GPC bound *Mtb* UgpB determined after this assay (Chapter 5, 5.3.6) this is likely to be the case. Other lipids that showed weaker binding were the di-

phosphorylated inositol phospholipids, phosphatidyl-inositol-3, 4-diphosphate (PtdIns(3, 4) P_2), phosphatidyl-inositol-3, 5-diphosphate (PtdIns(3, 5) P_2) and phosphatidyl-inositol-4, 5-diphosphate (PtdIns(4, 5) P_2) however these lipids showed lower binding affinities compared to the mono phosphorylated inositol phospholipids. As the tri-phosphorylated phosphatidyl-inositol-3, 4, 5-triphosphate (PtdIns(3, 4, 5)P_3) did not show binding it may suggest that size of the lipid head-group is a key determinant for binding. In hindsight overall this assay may not be the best to investigate lipid head-group binding based on the structures of these lipids and what we know about the binding recognition from the *Mtb* UgpB structure (Chapter 5, 5.3.6).

4.3.8 *Mtb* UgpB phosphoinositide array

To investigate the phospholipid interactions further a phosphoinositide array was tested using the same method as described previously (4.3.7) (Chapter 2, 2.5.9). The phosphoinositide array contains different concentrations (100-1.56 pmol per spot) of the inositol phospholipids and therefore concentration dependent binding could indicate the particular lipid binding preference. The results of the analysis showed that PtdIns(5)P had the highest binding as binding was found in the 12.5 pmol spot whereas the second highest binding was for PtdIns(4)P and the di-phosphorylated inositol phospholipid PtdIns(3, 5)P_2 as lower binding was also observed in the 12.5 pmol spot (Figure 4.12). The structures of the compounds that indicated strongest binding are shown (Figure 4.13) These results suggested that *Mtb* UgpB binding had a preference for phosphoinositide lipids with a phosphate at position 5 and 4 of the inositol ring.

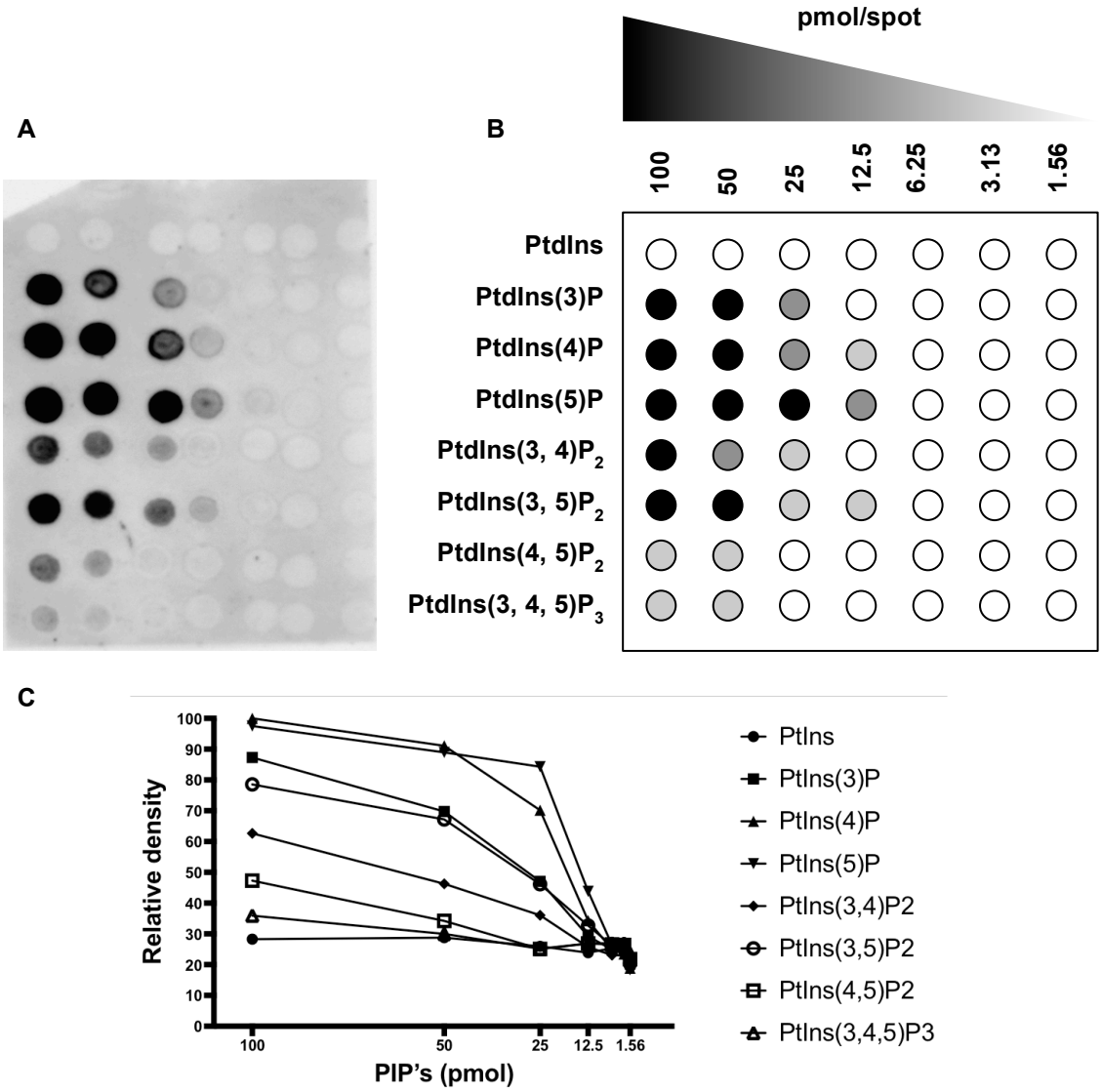


Figure 4.12: Phosphoinositide array *Mtb* UgpB. A) Phosphoinositide array probed with *Mtb* UgpB and anti-His/anti HRP secondary and ECL reagent. B) Schematic of phosphoinositide array, black indicates strongest binding, grey weaker binding and white no binding. C) Quantitation of phosphoinositide array binding calculated in ImageJ (n=1).

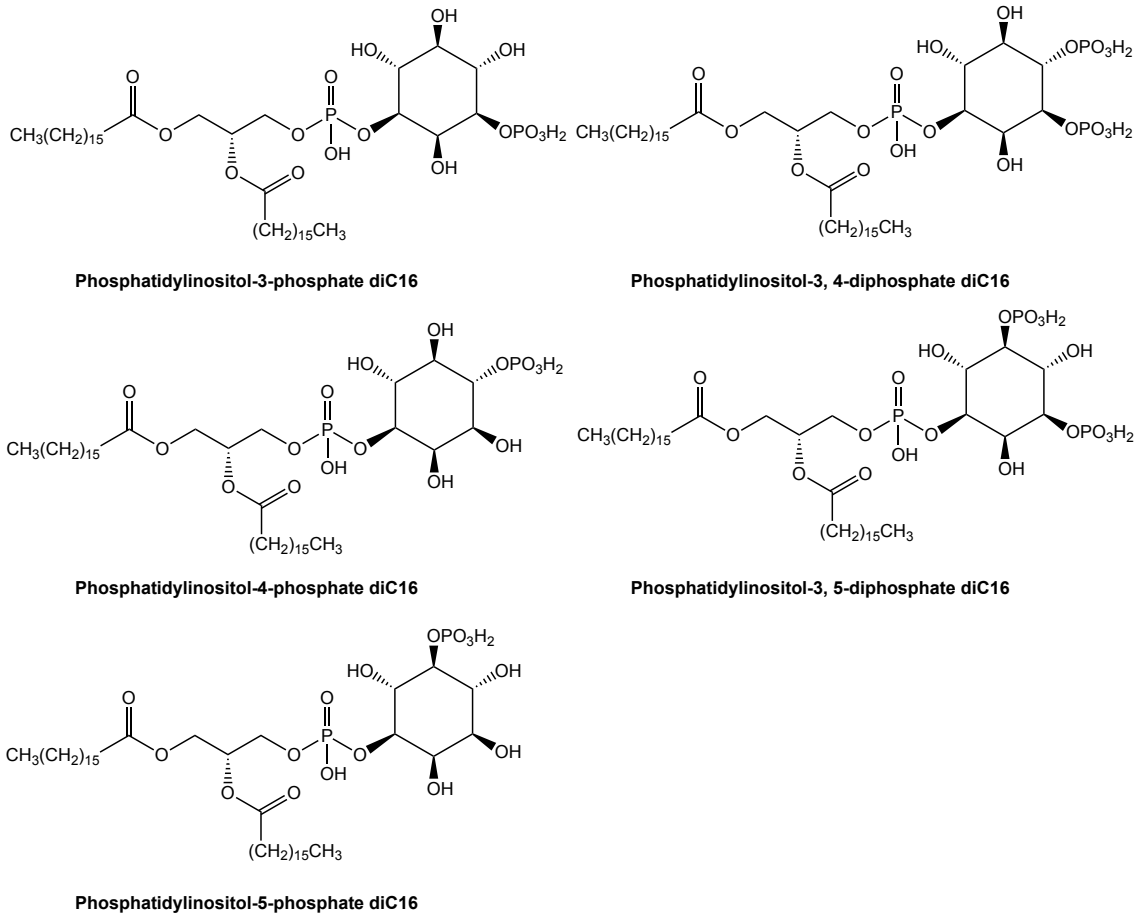


Figure 4.13: Structures of phospholipids that indicated the strongest binding to *Mtb* UgpB by phosphoinositide array,

4.3.9 *Mtb* UgpB sphingolipid strips

Another strip containing sphingolipids was tested (Figure 4.14) using the same method as described previously (4.3.7) (Chapter 2, 2.5.9). In this strip binding was only observed for sulfatide, a structurally different lipid to phospholipids that is a sulphur containing glycolipid (Figure 4.15). Sulfatide is derived from a ceramide lipid however it has a sulphate group that replaces the 3-hydroxyl of galactose. Interestingly its head-group does not have a phosphate group however sulfatides are important components of the *Mycobacteria* cell wall (Goren et al., 1976).

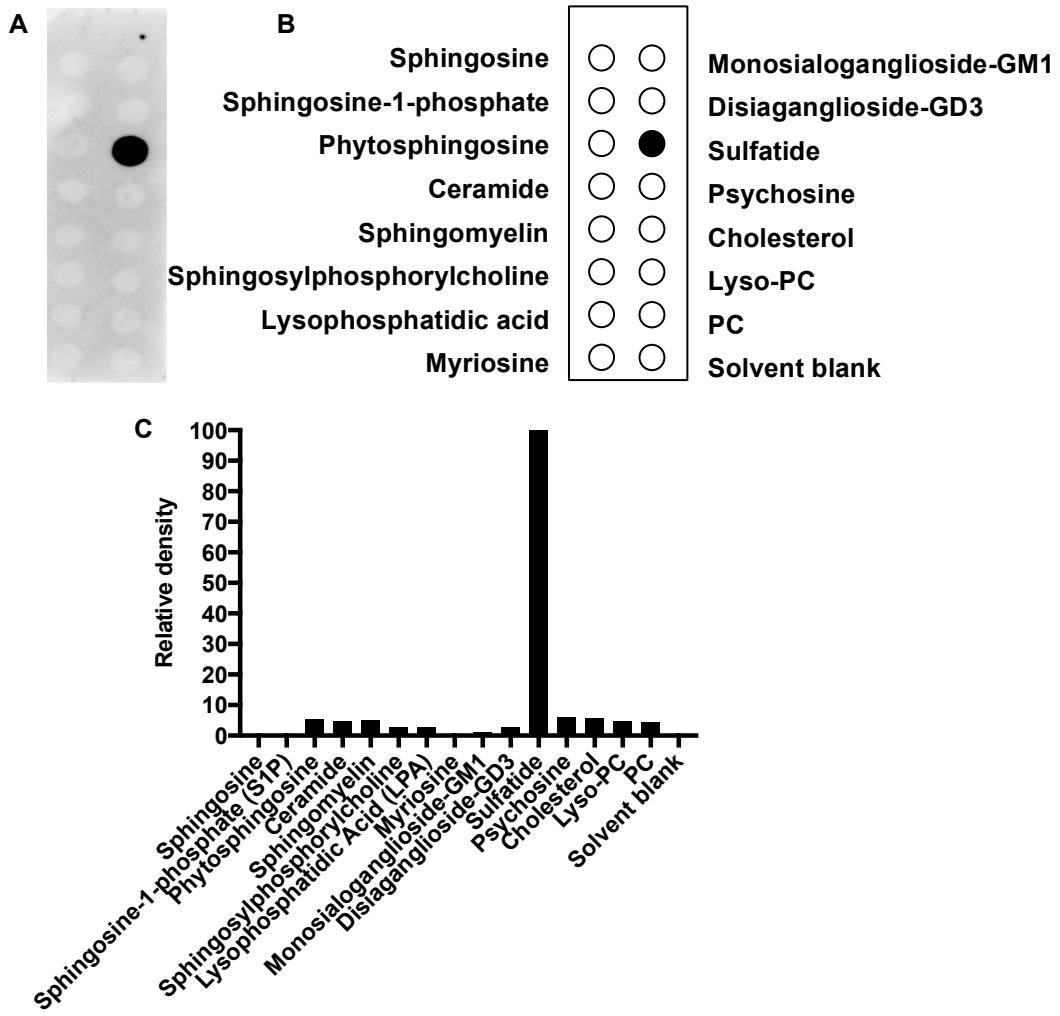
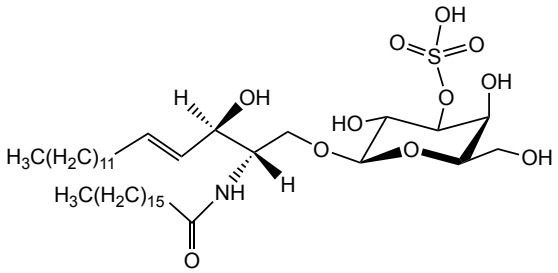


Figure 4.14: Sphingolipid strip *Mtb* UgpB. A) Sphingolipid strip probed with *Mtb* UgpB and anti-His/anti HRP secondary and ECL reagent. B) Schematic of sphingolipid strip, black indicates strongest binding, grey weaker binding and white no binding. C) Quantitation of phospholipid strip binding calculated in ImageJ (n=1).



Sulfatide

Figure 4.15: Structure of sulfatide that indicated binding to *Mtb* UgpB by sphingolipid strip.

4.3.10 Microarray screen of ligand binding to *Mtb* UgpB

To carry out a more extensive screen of ligand binding to *Mtb* UgpB a microarray analysis was carried out. The microarray screened for binding of 885 sequence-defined lipid-linked glycan probes to *Mtb* UgpB. Our collaborators Dr Lisete Silva and Dr Yan Liu at Imperial College London carried out the microarray analyses of *Mtb* UgpB.

Two different assay conditions were used (Chapter 2, 2.5.10), either with or without pre-complexation of the hexa-histidine tagged *Mtb* UgpB protein and the primary mouse anti-His and secondary biotinylated anti-mouse IgG detection antibodies before the overlay. The pre-complexation procedure allowed for artificial oligomerization of the hexa-histidine tagged *Mtb* UgpB and can enhance avidities of binding of monomeric proteins, or proteins with low binding affinities (Otto et al., 2011). For the non-precomplexed experiment, *Mtb* UgpB protein was overlaid followed by detection with primary monoclonal anti-polyHistidine antibody and secondary biotinylated anti-mouse IgG antibody. For the pre-complex experiment, protein-antibody complexes were prepared by pre-incubating primary antibody and secondary antibody, followed by addition of *Mtb* UgpB protein. For both the non-precomplexed and precomplexed protein, binding was detected with streptavidin-Alexa Fluor 647. However the microarray results found only non-specific binding to the probes and no significant binding signals were detected (Figure 4.16A and 4.16B).

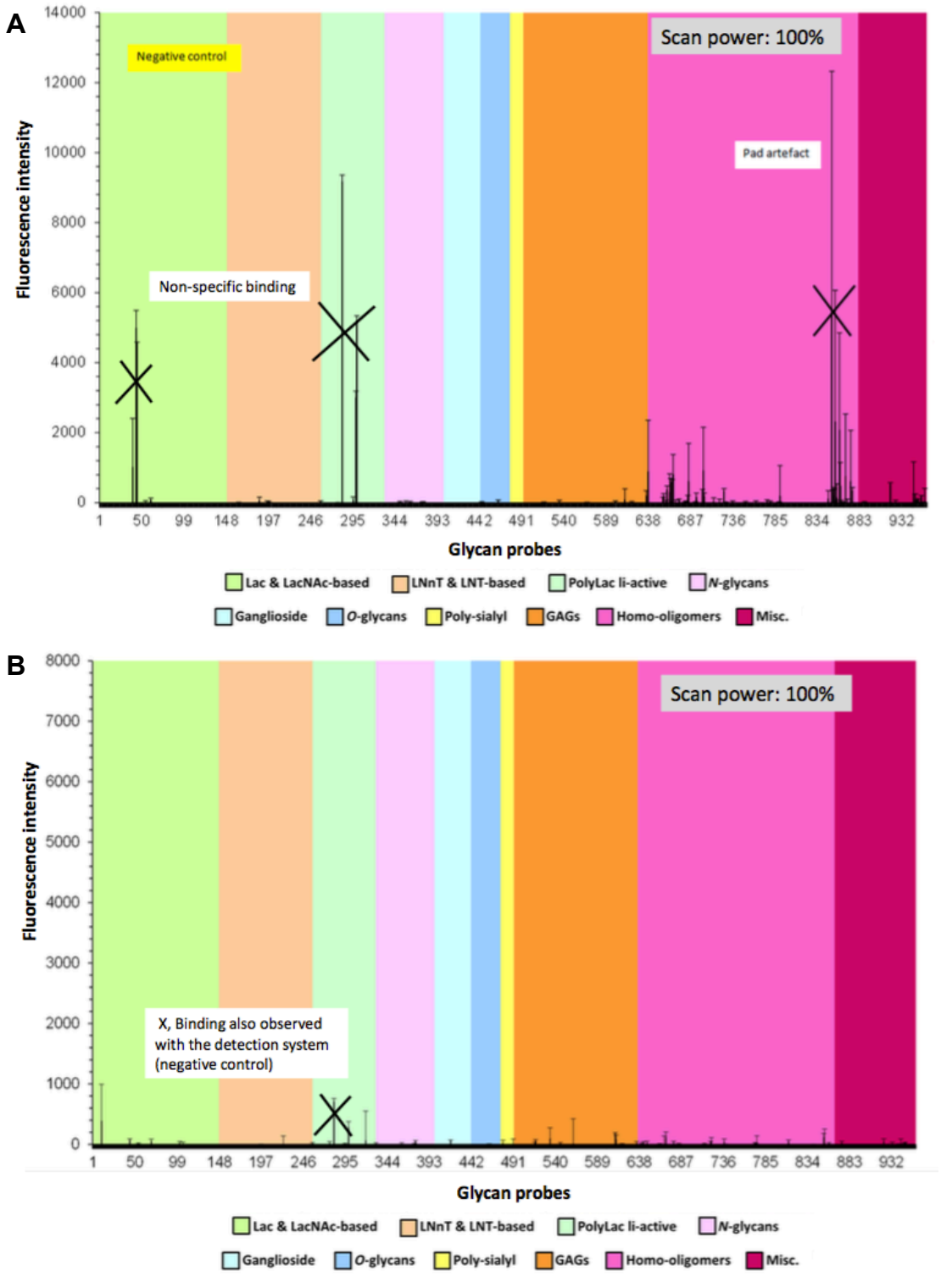


Figure 4.16: Microarray analyses of *Mtb* UgpB and 885 sequence-define lipid-linked glycan probes. A) Precomplexed *Mtb* UgpB. 25 (μg/mL) precomplexed 1:2:2 (w/w/w). B) *Mtb* UgpB without precomplexing 100 (μg/mL). Crosses refer to false positive signals.

4.3.11 BioLayer interferometry of *Mtb* UgpB

We wanted to establish a high throughput screen to investigate if any other substrates or potential inhibitors bind to *Mtb* UgpB. Therefore we decided to evaluate the use of BioLayer interferometry (BLI) because BLI has previously been used for drug screening (Wartchow et al., 2011). BLI is capable of detecting small molecule interactions with a sensitivity down to a molecular weight of >150 Da (ForteBio) (Concepcion et al., 2009). In this study we chose to immobilise *Mtb* UgpB to Octet NHS reactive BLI sensors (ForteBio) that react with free amines. As a result the buffer for *Mtb* UgpB could not contain free amines and therefore *Mtb* UgpB was dialysed from HEPES buffer into three different buffers: buffer A (25 mM sodium acetate, 150 mM NaCl, 10 % glycerol pH 5.5) was selected because the commercial recommendations (ForteBio) for NHS reactive BLI sensors were to use 25 mM sodium acetate at pH 5.5. Buffer B (25 mM NaH_2PO_4, 150 mM NaCl, 10 % glycerol pH 7.0) was selected because we wanted to keep the pH 7.0 and remove HEPES because it is a tertiary amine and may interfere with loading to the NHS reactive BLI sensors. Buffer C (25 mM HEPES, 150 mM NaCl, 10 % glycerol pH 7) was selected because we wanted to see if the original protein buffer could be used to measure binding. A final concentration of 20 μg/mL of *Mtb* UgpB was used to immobilise the protein onto the sensors for each buffer tested.

The sensorogram steps are shown (Figure 4.17) (Chapter 2, 2.5.11). Before immobilising *Mtb* UgpB to the sensor, the carboxylic acid surface was activated by reacting with 1-Ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride (EDC) (10 mM) and N-hydroxysulfosuccinimide (NHS) (20 mM) to generate sulfo-NHS esters, that are reactive with primary amines located in the protein. After immobilising *Mtb* UgpB (20 μg/mL) the unreacted sulfo-NHS esters were quenched by addition of ethanolamine (1 M), that has free amines, and the sensor was washed in buffer to remove any excess ethanolamine. Association of GPC to *Mtb* UgpB was then tested at 1 mM, this is approximately 36 times greater than the K_d of 27.3 μM determined by a previous study (Jiang et al., 2014).

We found that *Mtb* UgpB was successfully immobilised onto the sensor because an increase in the binding signal was observed in the 360-960 s UgpB loading step (Figure 4.17). Of the buffers tested (A-C) the optimal buffer was found to be a variation of the commercially recommended sodium acetate pH 5.5 (buffer A) because this buffer showed the largest increase in signal of approximately 1.0 nm in the loading step (Figure 4.17). This was because the slightly acidic pH 5.5 of buffer A was necessary for

optimal protein binding to the activated carboxylic acid surface of the BLI sensor. Minimal dissociation of *Mtb* UgpB was identified in the 960-1260 s ethanolamine (1 M) quenching step and the 1260-1380 s buffer baseline was stable indicating that the protein was stably bound (Figure 4.17).

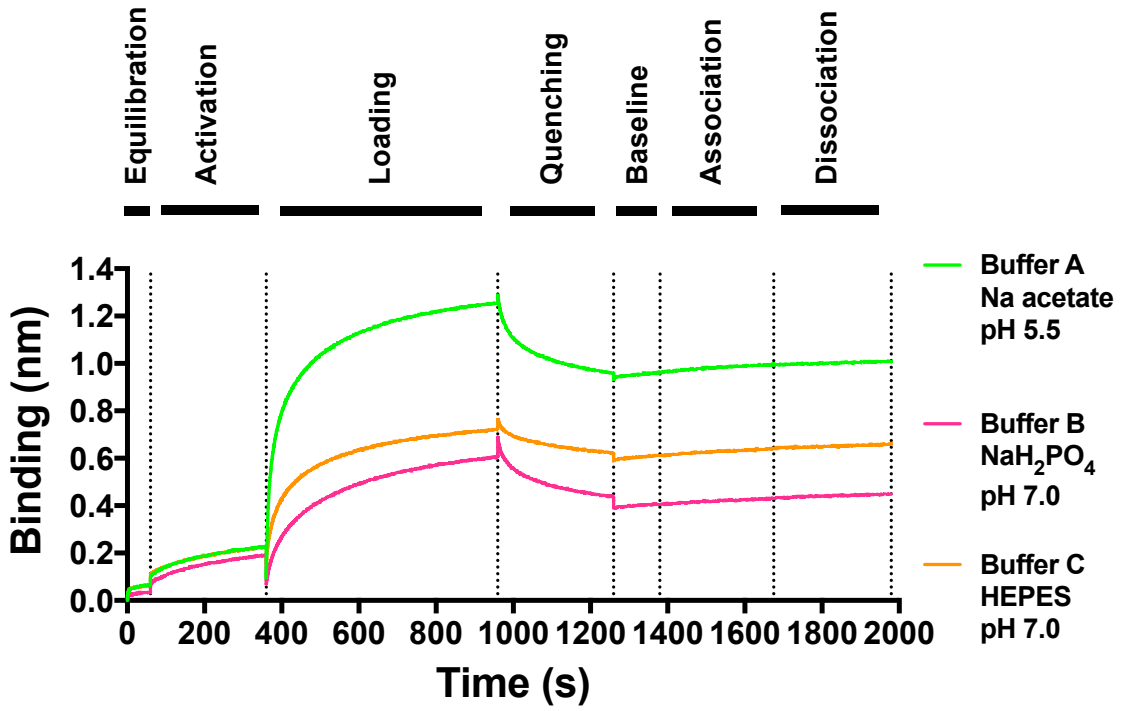


Figure 4.17: BioLayer interferometry sensorogram of *Mtb* UgpB with selected buffers. Green line refers to buffer A. Orange line refers to buffer B. Magenta line refers to buffer C. 0-60 s dH_2O equilibration, 60-360 s 1-Ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride (20 mM) (EDC)/NHS (10 mM) activation, 360-960 s UgpB loading, 960-1260 s ethanolamine (1M) quenching of the excess sulfo-NHS esters, 1260-1380 s buffer baseline, 1380-1675 s GPC (1 mM) association and 1675-1980 s GPC dissociation.

However association to GPC was not observed because there was no increase in the binding signal during the 1380-1675 s GPC (1 mM) association step or any decrease in the binding signal in the 1675-1980 s GPC dissociation step (Figure 4.17). The molecular weight detection limit for BLI is >150 Da and GPC approaches this limit, as its molecular weight is 257.22 Da, which may be why no association was observed. However it's also possible that ethanolamine (1 M) added during the 960-1260 s sulfo-NHS ester quenching step could be binding to *Mtb* UgpB and preventing GPC from binding to the binding site and hence why no signal was observed in the association step. Also as the buffer during the GPC association step was pH 5.5 this may have affected binding.

Although association of GPC to *Mtb* UgpB couldn't be detected using the conditions we tested, we wanted to see if the concentration of *Mtb* UgpB immobilised onto the sensor

could be lowered. As this would minimise the amount of protein used per assay. Therefore we tested different concentrations of *Mtb* UgpB: 5 μg/mL, 10 μg/mL and 20 μg/mL in buffer A (Figure 4.18). We found that 5 μg/mL of *Mtb* UgpB was sufficient for immobilisation as an increase in the binding signal to approximately 1.0 nm was observed in the 360-960 s UgpB loading step (Figure 4.18)

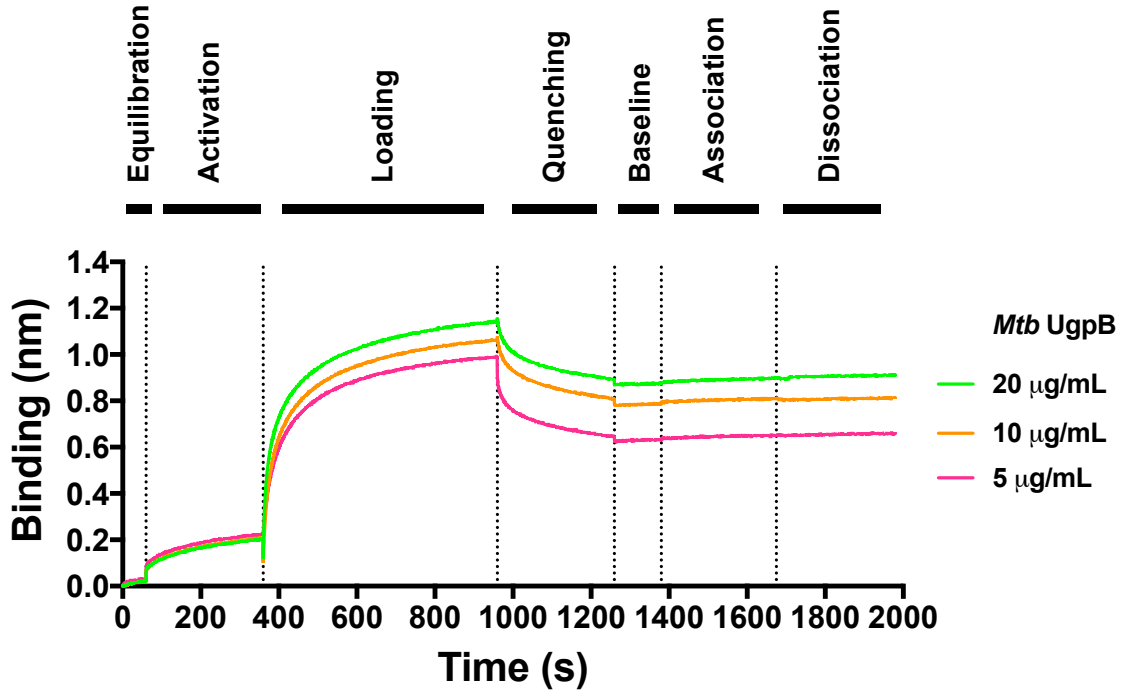


Figure 4.18: BioLayer interferometry sensorogram of *Mtb* UgpB loading optimisation. Green line refers to 20 μg/mL protein. Orange line refers to 10 μg/mL protein. Magenta line refers to 5 μg/mL. 0-60 s dH_2O equilibration, 60-360 s 1-Ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride (20 mM) (EDC)/NHS (10 mM) activation, 360-960 s UgpB loading, 960-1260 s ethanolamine (1 M) quenching of the excess sulfo-NHS esters, 1260-1380 s buffer baseline, 1380-1675 s GPC (1 mM) association and 1675-1980 s GPC dissociation.

In the future it may be better to biotinylate *Mtb* UgpB and load it onto super streptavidin biosensors, which are designed for the analysis of small molecule interactions (ForteBio). In particular super strepatavidin biosensors would also avoid using the coupling agent ethanolamine that is potentially remaining and binding to *Mtb* UgpB. Additionally, super streptavidin biosensors are compatible with buffers at a range of different pH. As pH may influence binding to GPC, the original purification buffer could be used instead. However, instead of optimising BLI for ligand screening we chose to investigate MST to determine binding affinities.

4.3.12 Binding affinity analysis of *Mtb* UgpB with GPC by microscale thermophoresis

To investigate the interactions of *Mtb* UgpB with the potential ligands identified from the thermal shift analysis (4.3.2) and lipid strips (4.3.7), MST was used.

The *Mtb* UgpB protein was labelled with RED-Tris-NTA dye that binds to the C-terminal hexa-His tag. Full details of the method are described (Chapter 2, 2.5.12). *Mtb* UgpB was used at a final concentration of 50 nM in the assay and binding to compounds in the concentration range (0-0.5 M) was measured using the MonoLith NT.115 instrument with medium laser power and 40 % LED power settings. The binding affinities were calculated and the K_d model was fitted in the built-in MST analysis software. The calculated K_d model was plotted in GraphPad Prism V8. All experiments were calculated in triplicate.

Initial binding of *Mtb* UgpB to GPC was tested and a high binding affinity of 3.6 μM ± 0.5 μM was determined (Figure 4.19) (Table 4.1). This result was comparable with a previous study that showed GPC had a binding affinity of 27.3 μM by ITC (Jiang et al., 2014). However, as expected based on the thermal shift assay (4.3.5) and the structure of GPC bound *Mtb* UgpB determined later (Chapter 5, 5.3.6), binding to G3P or phosphocholine could not be detected by MST (data not shown).

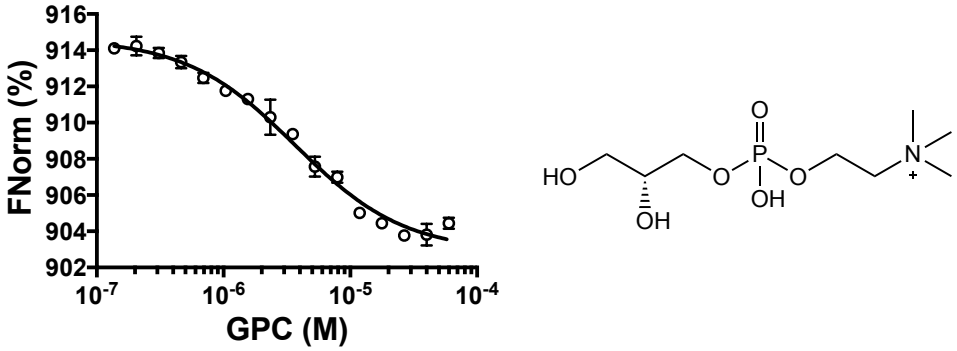


Figure 4.19: Microscale thermophoresis *Mtb* UgpB and glycerol-3-phosphocholine. FNorm (%) is the normalised fluorescence signal of the change in MST signal. Error bars represent SD from at least three independent experiments.

4.3.13 Binding affinity analysis of *Mtb* UgpB mutant derivatives by microscale thermophoresis

Residues identified from the X-ray structure of the GPC bound *Mtb* UgpB binding site (Chapter 5, 5.3.6) were each mutated to alanine and the site-directed mutants of *Mtb* UgpB were successfully purified (Chapter 3, 3.3.9). MST was used to test the mutant derivatives of *Mtb* UgpB ability to bind to GPC. All of the mutants were prepared for

MST analyses as described previously (4.3.13) and (Chapter 2, 2.5.12). MST analysis indicated that six of the mutants tested (Table 4.1) did not bind to GPC. Two of the mutants, Ser153Ala and Leu205Ala had a lower binding affinity compared to apo-*Mtb* Leu205Ala was found to have a binding affinity of 161.7 μM and Ser153Ala was found to have a binding affinity of 309.8 μM, approximately 45-fold and 85-fold lower compared to apo-*Mtb* UgpB (Table 4.1).

Table 4.1: Summary of *Mtb* UgpB MST binding affinity data.

Protein	Substrate	K_d (μM)
Mtb UgpB	GPC	3.6 ± 0.5
Mtb UgpB	GPC*	27.3
Mtb UgpB Y78A	GPC	-
Mtb UgpB D102A	GPC	-
Mtb UgpB S153A	GPC	309.8 ± 56.1
Mtb UgpB L205A	GPC	161.7 ± 15.9
Mtb UgpB W208A	GPC	-
Mtb UgpB S272A	GPC	-
Mtb UgpB Y345A	GPC	-
Mtb UgpB R385A	GPC	-

(-) = no binding detected, ± SD from three independent experiments. GPC: glycerol-3-phosphocholine. GPC* is the binding affinity determined by ITC (Jiang et al., 2014).

4.3.14 Binding of alternative glycerophosphodiesters to *Mtb* UgpB

The GPC bound X-ray structure of *Mtb* UgpB (Chapter 5, 5.3.6) did not indicate any *Mtb* UgpB binding site interactions with the choline head-group of GPC. However it was possible that previously untested and physiologically relevant glycerophosphodiesters could also bind to *Mtb* UgpB. Two of these glycerophosphodiesters were commercially available: glycerophosphoinositol-4-phosphate (GPI(4)P) and glycerophosphoinositol (GPI) however glycerophosphoethanolamine (GPE) and glycerophosphoserine (GPS) were not commercially available and had to be produced. To do this the following synthetic routes were first attempted (4.4).

4.4 Glycerophosphodiester synthesis

4.4.1 Glycerophosphodiester synthetic route

4.4.1.1 Glycerophosphodiester synthetic route

A synthetic route was attempted to produce the glycerophosphodiesters: glycerophosphoethanolamine **(1)** (GPE) and glycerophosphoserine **(2)** (GPS). This was based on two previous studies (R.-R. Tang et al., 2005) (Gagnon et al., 2017) that successfully synthesised glycerophosphodiesters from the respective building blocks **(3)**, **(4)** and **(5)** (Figure 4.20). Retrosynthetic analysis (Figure 4.20) indicates that the products **(1)** and **(2)** are split into three components: a glycerol group **(3)**, phosphate group **(4)** and head-group **(5)** containing either ethanolamine or serine.

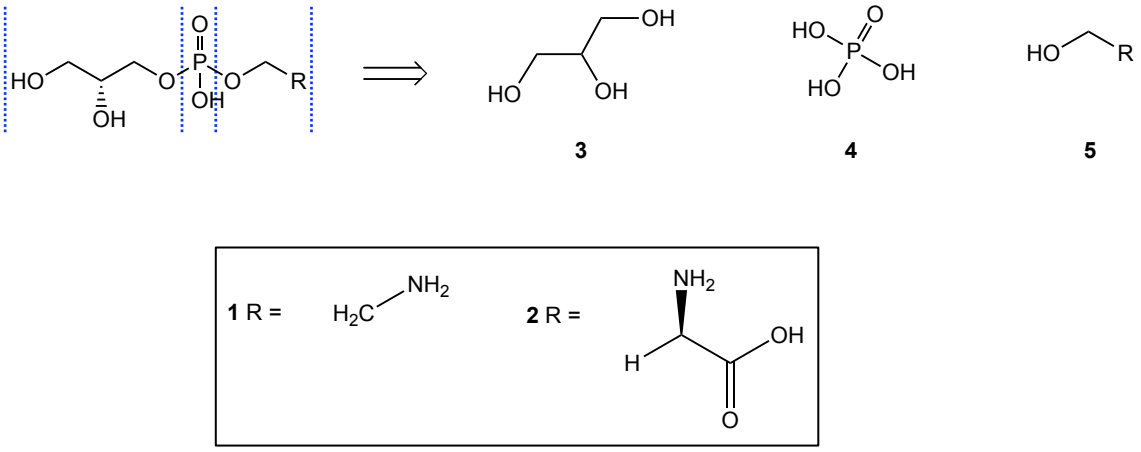
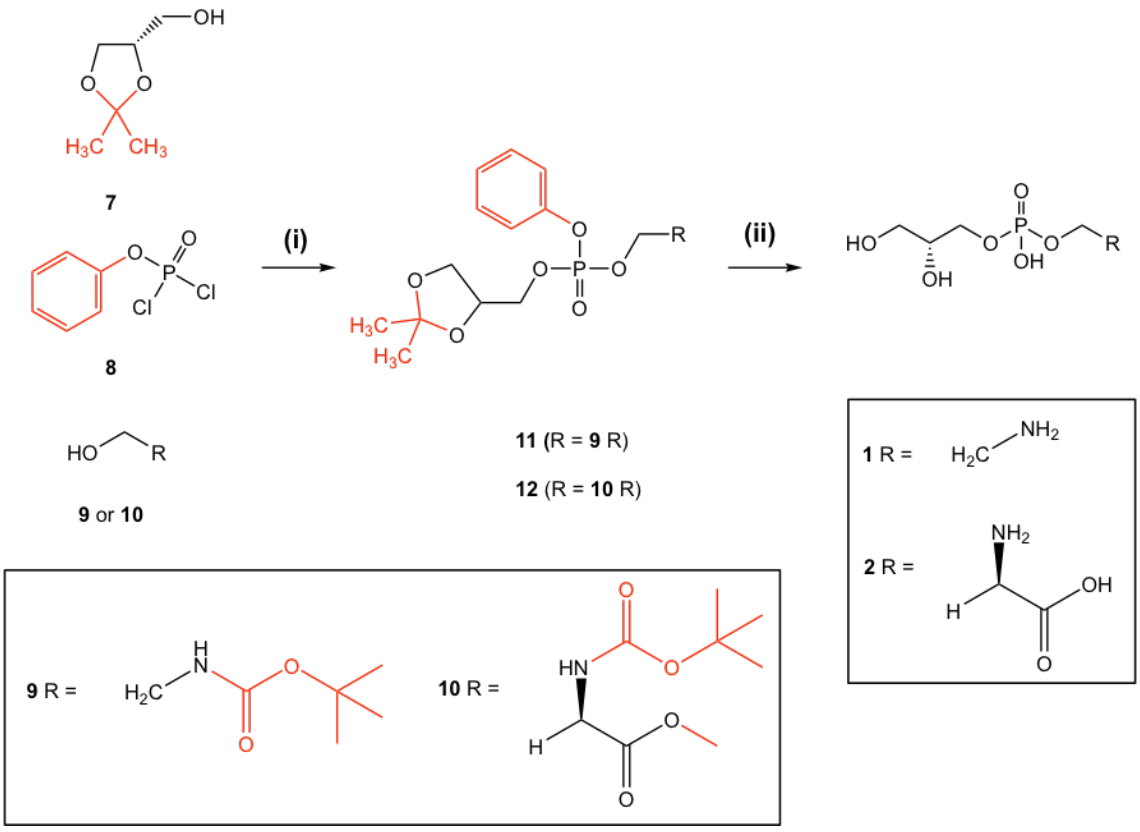


Figure 4.20: Retrosynthetic analysis for the production of glycerophosphoethanolamine (GPE) and glycerophosphoserine (GPS).

The precursors (Scheme 4.1) **(7)** and **(9)** or **(10)** could react via their un-protected hydroxyl groups with the phosphate group of precursor **(8).** As the other hydroxyl groups of **(7)** and **(9)** or **(10)** were protected the reaction could be directed to produce the protected glycerophosphodiesters **(11)** or **(12)**. Therefore it was necessary to use protected versions of **(3)**, **(4)** and **(5)**. The protected reagents included: (S)-isopropylideneglycerol **(7)**, the glycerol precursor containing an acetonide protective group so that only a single hydroxyl was reactive. Phenyldichlorophosphate **(8)**, the phosphate precursor containing a protective phenyl group and two reactive chloride atoms. *N*-Boc-ethanolamine **(9)** or *N*-Boc-L-serine methyl ester **(10)** are the ethanolamine and L-serine precursors containing an *N*-Boc protective group and reactive hydroxyl group. The reaction of **(7)** and **(8)** with **(9)** or **(10)** would produce the intermediates: protected glycerophosphoethanolamine **(11)** or protected glycerophosphoserine **(12)**. Once this protected product was purified, removal of the protective phenyl with a base such as NaOH, and removal of the *N*-Boc group and

acetonide group with an acid such as TFA would produce the products glycerophosphoethanolamine (GPE) **(1)** and glycerophosphoserine (GPS) **(2)**. The proposed overall reaction scheme is described (Scheme 4.1).



Scheme 4.1: Proposed overall reaction scheme for the synthesis of glycerophosphoethanolamine (GPE) and glycerophosphoserine (GPS). The different strategies attempted are described (Table 4.2). Red indicates the protecting groups. (i) refers to protected glycerophosphodiester synthesis. (ii) refers to deprotection of the protected glycerophosphodiester to produce the product.

4.4.1.2 Glycerophosphodiester synthesis strategies

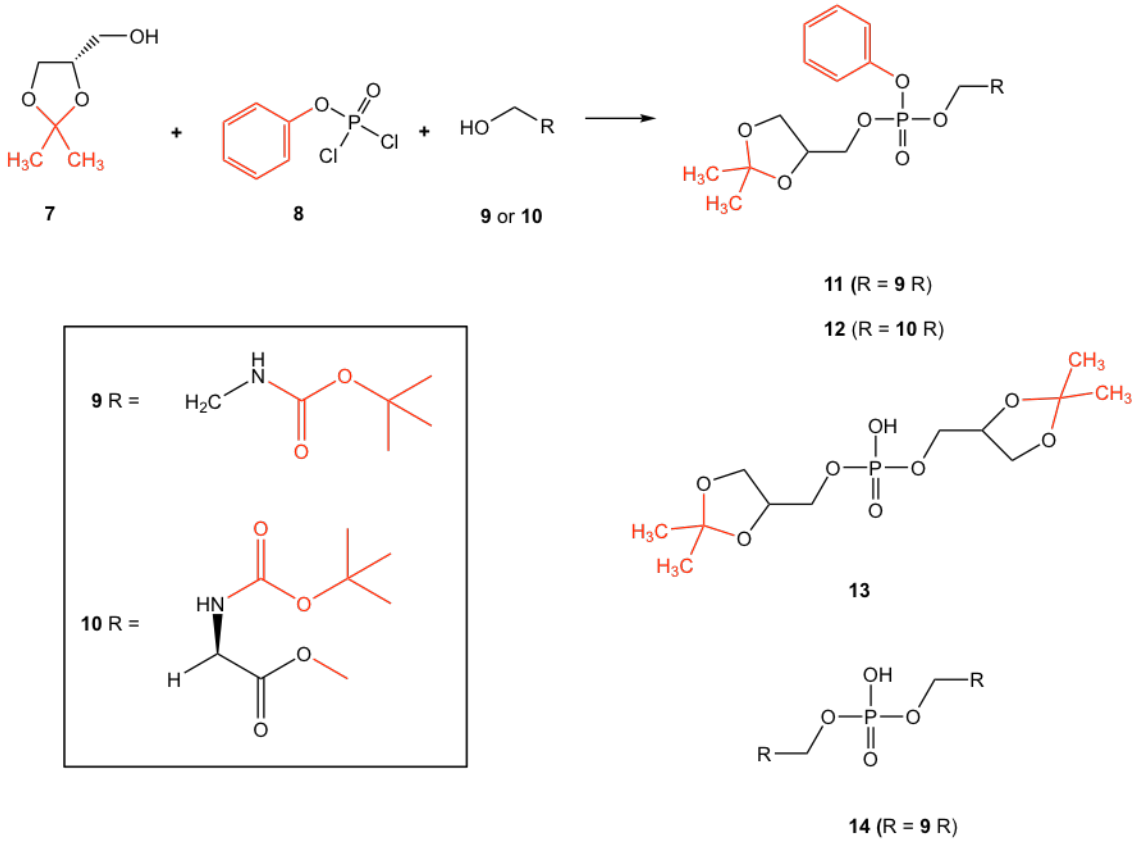
The first step was to produce the protected glycerophosphodiesters **(11)** and **(12)** as detailed in the overall scheme (Scheme 4.1). Three different strategies were attempted to produce the protected glycerophosphodiesters **(11)** and **(12)** and these are detailed (Table 4.2). The strategies differed based on the order of reagent addition and the type of base added (Table 4.2).

Table 4.2: Glycerophosphodiester synthesis strategies.

Strategy	First addition	Second addition	Base
1	Isopropylidene glycerol (1.0 eq) $PhOPOCl_2$ (1.0 eq) Solvent: Toluene (5 mL)	*N*-Boc-ethanolamine (1.0 eq) Or *N*-Boc-L-serine methyl ester (1.0 eq) Solvent: Chloroform (5 mL)	Pyridine (1.5 eq first addition followed by 4.0 eq second addition)
2	Isopropylidene glycerol (1.0 eq) $PhOPOCl_2$ (1.0 eq) Solvent: Toluene (5 mL)	*N*-Boc-ethanolamine (1.0 eq) Or *N*-Boc-L-serine methyl (1.0 eq) ester Solvent: Chloroform (5 mL)	Triethylamine (1.5 eq first addition followed by 4.0 eq second addition)
3	$PhOPOCl_2$ (1.0 eq) *N*-Boc-ethanolamine (1.0 eq) Or *N*-Boc-L-serine methyl ester (1.0 eq) Solvent: Chloroform (5 mL)	Isopropylidene glycerol Solvent: Toluene (5 mL)	Triethylamine (1.5 eq first addition followed by 4.0 eq second addition)

4.4.1.3 Strategy 1 synthesis

The reagents and scheme for this strategy are described (Table 4.2) (Scheme 4.2). In order to attempt to synthesise the protected intermediate products **(11)** and **(12)**, we reacted **(7)** with **(8)** followed by the addition of either **(9)** or **(10)** (Scheme 4.2) (Table 4.2)**.** The reactions were carried out under N_2 and with stirring. During the first reagent addition: **(7)** reacted with **(8)**, we found it was essential to keep the temperature at -20 °C to avoid the protected glycerol **(7)** reacting twice with the protected phosphate **(8)**. The reagents were allowed to reach room temperature over 6 hours. For the second reagent addition: **(9)** or **(10)** were added and the reagents were allowed to react at room temperature overnight.



Scheme 4.2: Strategy 1 and 2 scheme for the synthesis of glycerophosphoethanolamine (GPE) and glycerophosphoserine (GPS) using either pyridine or triethylamine base. (7) was reacted with **(8)** first, followed by reacting with **(9)** or **(10)**.

The reaction mixture was worked up by washes with: 50 mL ethyl acetate, 70 mL 10 % HCl, 70 mL H_2O, 70 mL $NaHCO_3$ and 70 mL brine. The organic phase was collected, dried by addition of anhydrous $MgSO_4$ and vacuum filtered. The crude product in the organic phase was analysed by NMR and mass spectrometry.

In strategy 1 (Table 4.2), pyridine was used as the base. Strategy 1 did not produce much product, as by NMR the starting materials were not completely used therefore we decided to repeat the reaction in strategy 2 (Table 4.2) (Scheme 4.2) using triethylamine as the base instead of pyridine because triethylamine is a stronger base. Note that strategy 1 and 2 are identical except a different base is used.

4.4.1.4 Strategy 2 synthesis

The reaction was repeated as in (4.4.1.3) hence the scheme is identical (Scheme 4.2) however the base was changed from pyridine to triethylamine, which is a stronger base. Strategy 2 (Scheme 4.2) (Table 4.2) resulted in production of a mixture of the protected products GPE 454 m/z **(11)** and GPS 512 m/z **(12),** the 2x glycerol reacted by-product 424 m/z **(13)** and the 2x ethanolamine reacted by-product 483 m/z **(14)** as determined by mass spectrometry (Figure 4.21). The starting materials were

completely used as determined by mass spectrometry and NMR, however the ratio of the desired products **(11)** and **(12)** to by-products **(13)** and **(14)** wasn't particularly high therefore we decided change the order of reagent addition in strategy 3 (Table 4.2) (Scheme 4.3).

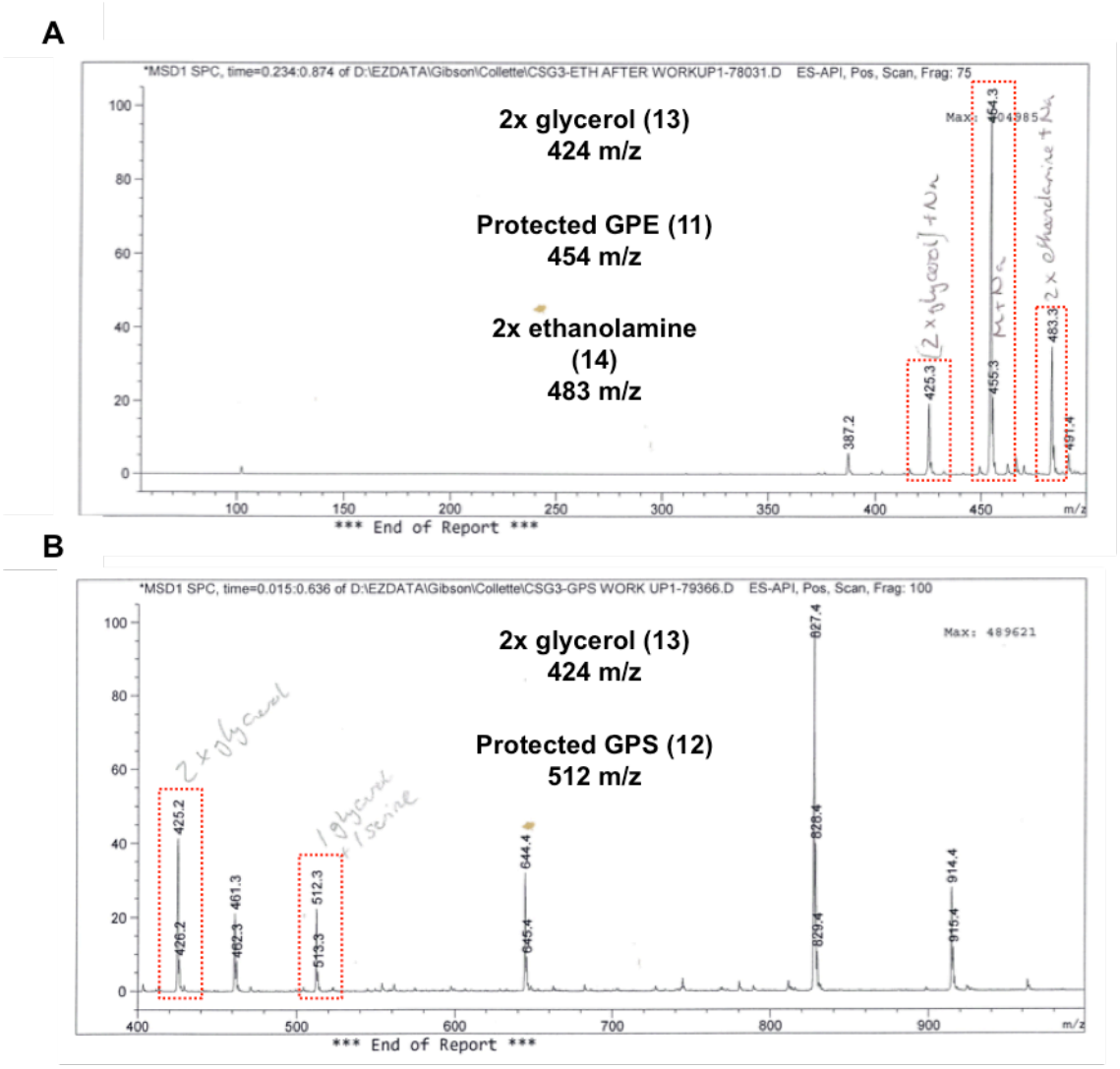
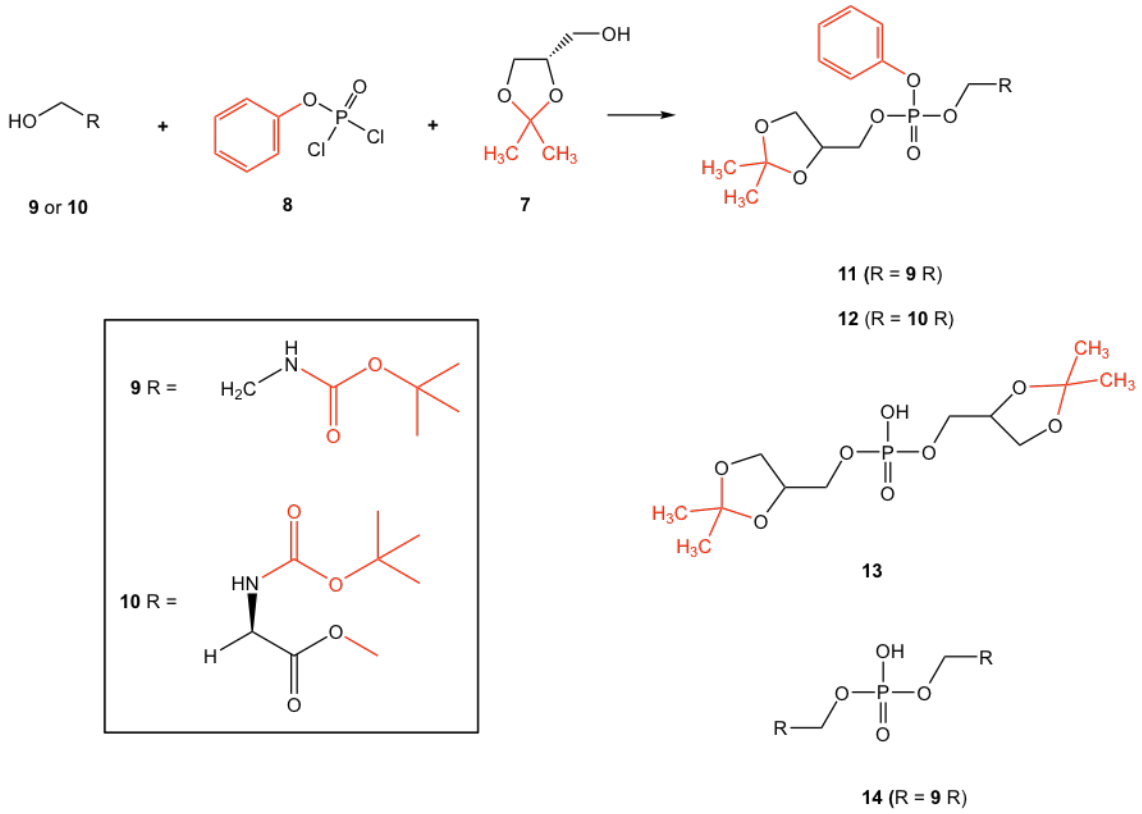


Figure 4.21: Products and by-products identified by mass spectrometry. A) Glycerophosphoethanolamine synthesis. B) Glycerophosphoserine synthesis.

4.4.1.5 Strategy 3 synthesis

In strategy 3 (Scheme 4.3) (Table 4.2) we changed the order of reagent addition and reacted **(9)** or **(10)** with **(8)** followed by the addition of **(7)** using triethylamine as the base and the same procedures as previously described (Table 4.2) (Scheme 4.3). The crude products were analysed by NMR and mass spectrometry and revealed a better ratio of the desired products **(11)** and **(12)** compared to by-products **(13)** and **(14)** therefore we decided to use strategy 3 moving forwards.



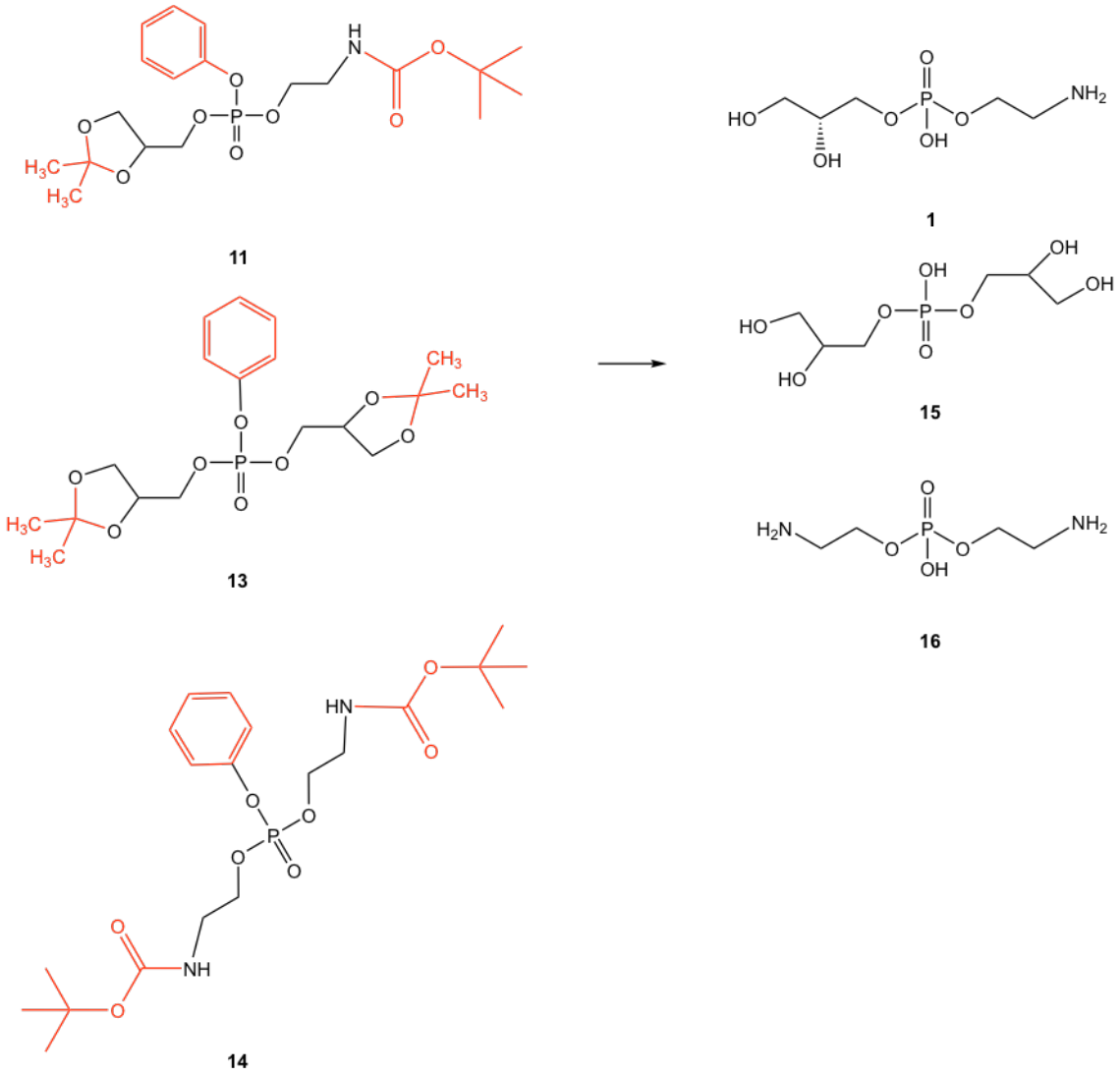
Scheme 4.3: Strategy 3 scheme for the synthesis of glycerophosphoethanolamine (GPE) and glycerophosphoserine (GPS) using triethylamine base and changing the order of reagent addition. (9) or (**10**) was reacted with **(8)** first, followed by reacting with **(7)**. The base used throughout was triethylamine.

To conclude, in strategy 2 and 3 (Table 4.2), there was complete conversion of starting materials **(7)**, **(8)** and **(9)** or **(10)** to the products **(11)** and **(12)** and by-products **(13)** and **(14)**. Based on the crude NMR and mass spectrometry we concluded that strategy 3 was optimal therefore the next step was to separate the crude protected product **(11)** and **(12)** from the by-products **(13)** and **(14)** (Scheme 4.3).

Column chromatography was used to separate the protected glycerophosphodiesters **(11)** and **(12)** from the by-products **(13)** and **(14)**. We used the solvent 2:1 (hexane/ethyl acetate v/v) and found that the protected GPS **(12)** did not separate well on the column despite trying several different polarities of solvent systems. This was potentially because the polarities of the GPS product **(12)** and the by-product **(13)** were more similar in the case for GPS therefore we were unable to separate them and did not do any further experiments for GPS. In contrast, GPE **(11)** separated better however NMR and mass spectrometry of GPE **(11)** after column chromatography indicated that there were still some by-products **(13)** and **(14)** present. The fractions containing the highest abundance of protected GPE **(11)** product were combined and we proceeded to remove the protecting groups (4.4.1.6).

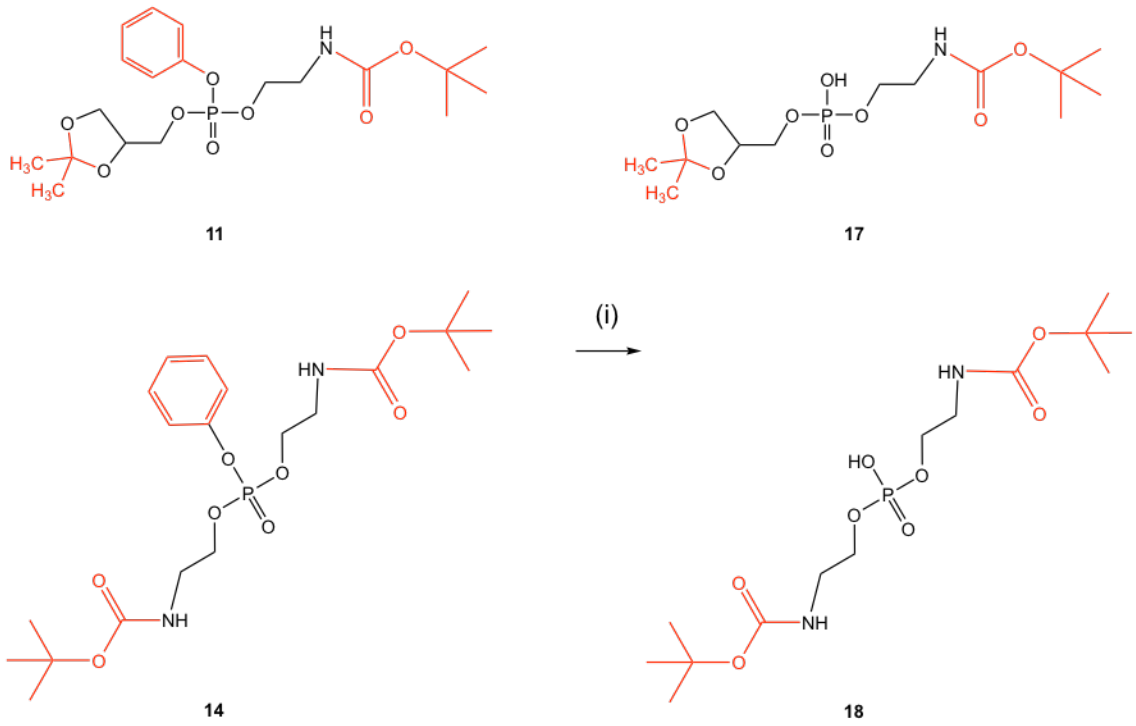
4.4.1.6 Deprotection of glycerophosphoethanolamine (GPE)

The overall deprotection scheme starting with the mixture of protected GPE product **(11)** and by-products **(13)** and **(14)** are described (Scheme 4.4). The aim of deprotection was to remove the acetonide group, phenyl group and *N*-Boc group to result in the correct GPE product **(1)** and a small amount of de-protected by-products **(15)** and **(16)**.



Scheme 4.4: Overall scheme for the deprotection of glycerophosphoethanolamine including by-products.

To deprotect the phenyl group of protected GPE **(11)** we first used the base NaOH (Scheme 4.5). We found it required the addition of 2.2 eq NaOH in tetrahydrofuran solvent with overnight incubation at room temperature to completely remove the phenyl group **(17)**. This was analysed using mass spectrometry (Figure 4.22) by the presence of phenyl removed GPE 354 m/z **(17)** and also some of the phenyl removed 2x ethanolamine by-product 383 m/z **(18)**.



Scheme 4.5: Removal of the phenyl group. (i) Addition of 2.2 eq NaOH in tetrahydrofuran incubated overnight at room temperature produced the phenyl-removed product **(17)** and phenyl removed 2x ethanolamine by-product **(18)**.

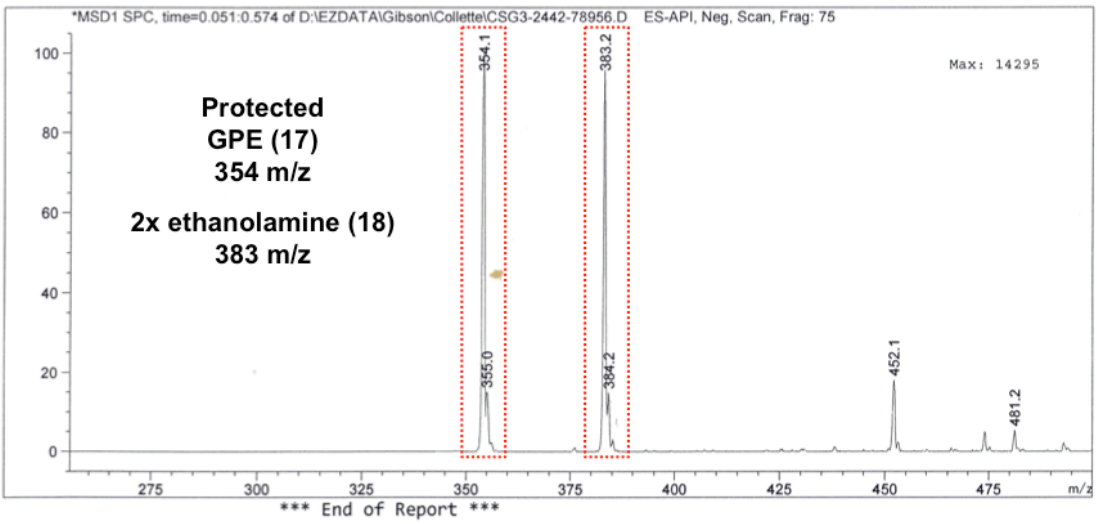
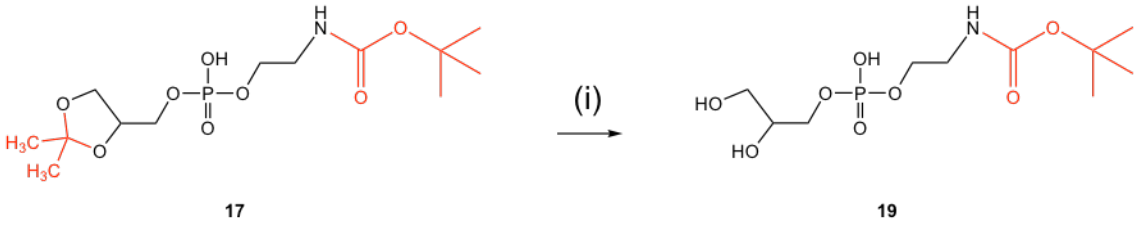


Figure 4.22: Mass spectrometry confirms removal of the phenyl group. Produced the phenyl-removed product **(17)** 354 m/z and phenyl removed 2x ethanolamine by-product **(18)** 383 m/z.

Next, removal of the acetonide and *N*-Boc groups from **(17)** required addition of acid (Scheme 4.6). We first added 5.0 eq TFA in tetrahydrofuran solvent however after overnight incubation at room temperature and mass spectrometry analysis (Figure 4.23) it revealed two major peaks at 314 m/z **(19)** and (383 m/z) **(18)** suggesting the

phenyl and acetonide groups were successfully removed however the *N*-Boc group was still intact, resulting in compound **(19)**.



Scheme 4.6: Removal of the acetonide group. (i) addition of 5.0 eq TFA in tetrahydrofuran incubated overnight at room temperature produced the acetonide removed product **(19)**.

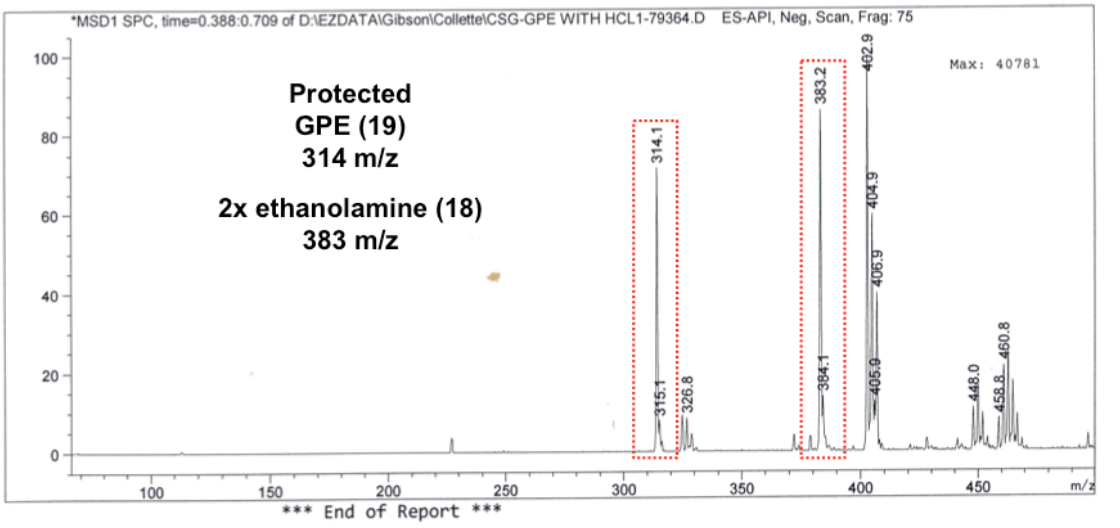
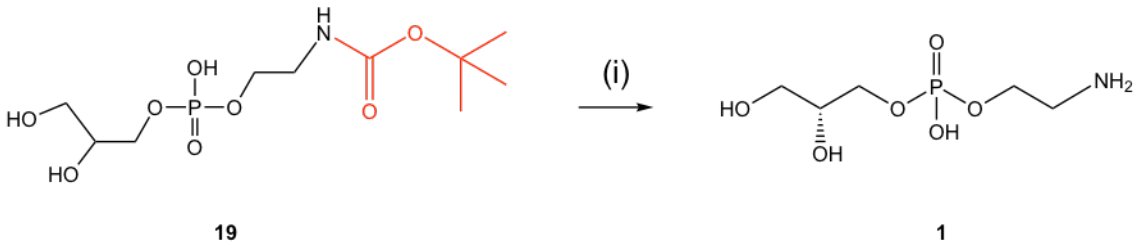


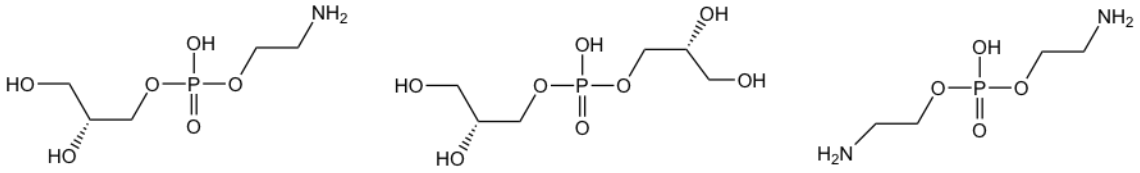
Figure 4.23: Mass spectrometry confirms removal of the acetonide group. Produced the acetonide-removed product **(19)** 314 m/z and also some of the phenyl removed 2x ethanolamine by-product **(18)** 383 m/z.

We finally found that addition of 6.0 eq NaOH in tetrahydrofuran solvent to compound **(11)** and overnight incubation at room temperature, followed by the addition of 10.0 eq TFA in dichloromethane and overnight incubation at room temperature successfully removed all of the protecting groups including *N*-Boc to produce the GPE product **(1)** (Scheme 4.7). However now as the protecting groups were removed the product **(1)** and by-products **(15)** and **(16)** (Figure 4.24) did not fly well in the mass spectrometer and therefore could not be detected by mass spectrometry.



19 1

Scheme 4.7: Removal of the *N*-Boc group. (i) addition of 10.0 eq TFA in dichloromethane incubated overnight at room temperature produced the acetonide removed product **(1)**.



1 15 16

Figure 4.24: Structures of deprotected glycerophosphoethanolamine product (1) and by-products (15) and (16).

After successfully removing the protective groups, the product **(1)** and by-products **(15)** and **(16)** were now soluble in H_2O, which enabled the organic soluble products to be removed by extraction. This resulted in a mixture of products **(1)**, by-products **(15)** and **(16)** and TFA inorganic salts. To remove the TFA inorganic salts an amberlite IR120 cation exchange resin was prepared. The deprotected products were loaded onto the column and eluted with dH_2O however only a small amount of product **(1)** and by-products **(15)** and **(16)** eluted. We thought the product **(1)** and by-products **(15)** and **(16)** were binding to the column due to the addition of acid during the deprotection conditions that may have protonated the NH_2 of ethanolamine to NH_3^+. Ammonia water was added to deprotonate the product **(1)** and by-product **(16)** however they could not be successfully eluted.

Instead we found that dissolving the deprotected products in a minimal amount of acetone successfully removed the TFA inorganic salts. Subsequent NMR analysis indicated that the acetone soluble products included the correct GPE product **(1)** however it also contained several contaminants most likely carried over due to the inability to completely separate the protected GPE **(11)** from by-products **(13)** and **(14)**.

To conclude we found that protected synthesis of glycerophosphodiesters from the individual building blocks was not very successful. We found we were losing product during removal of the inorganic impurities and we still had a mixture of by-products.

Therefore we decided to use a different method, starting with phospholipids and using either chemical or enzymatic methods to hydrolyse the lipid tails in order to produce the glycerophosphodiesters. After removal of the lipid tails, the lipids tails would be soluble in organic chloroform solvent whereas the head-group would be soluble in H_2O.

4.4.2 Base hydrolysis of phosphatidyl-L-serine

We attempted a chemical method based on the following studies that used the base NaOH to hydrolyse the fatty acid tails from phospholipids (Figure 4.25) (Kopp et al., 2010) (Nishihara et al., 1988). We started with a commercially available phospholipid, phosphatidyl-L-serine **(20)** extracted from cell membranes.

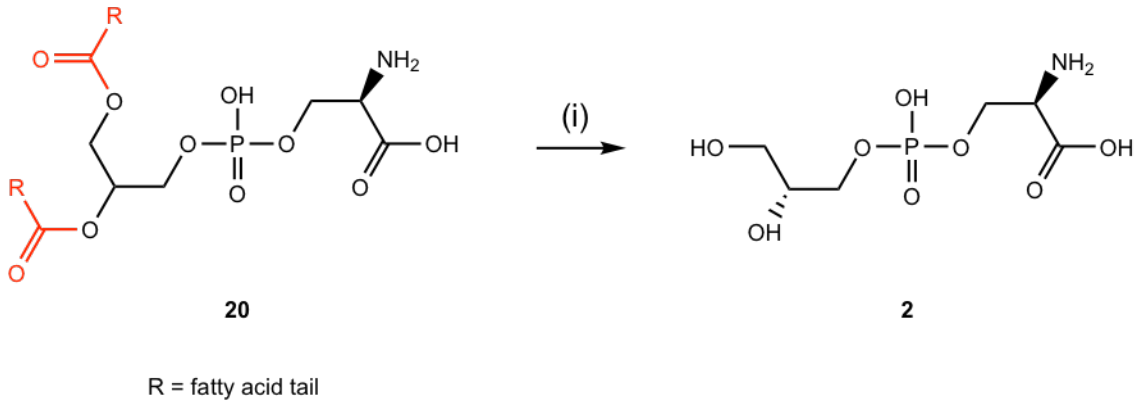


Figure 4.25: Base hydrolysis of phosphatidyl-L-serine (20) to produce glycerophosphoserine (1). (i) Addition of 2 M NaOH in 2:1 (chloroform/methanol v/v).

We dissolved 100 mg of phosphatidyl-L-serine **(20)** in 2:1 (chloroform/methanol v/v). 2 M NaOH was added and incubated for 2 hours at room temperature to hydrolyse the fatty acid tails. The reaction was neutralized by the addition of 2 M HCl and an organic aqueous extraction using chloroform and dH_2O was used to separate the fatty acid tails from the head-group. The aqueous fraction of the reaction was analysed by TLC in 5:4:2:1 (butan-1-ol/methanol/ammonia/dH_2O v/v/v/v) solvent. The TLC plates were stained with ninhydrin to detect the amine group of glycerophosphoserine **(2)** and also stained with sulphuric acid to compare any differences in staining. TLC analysis indicated at least two ninhydrin and sulphuric acid stained spots that may correspond to the glycerophosphoserine **(2)** and phosphoserine **(22)** products (Figure 4.26) therefore we analysed the samples by high-resolution mass spectrometry (Figure 4.27 and 4.28).

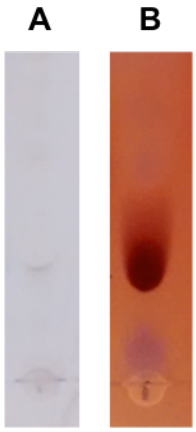


Figure 4.26: TLC analysis of the aqueous fraction of phosphatidyl-L-serine base hydrolysis products following aqueous/organic extraction. A) 10 % sulphuric acid in ethanol stain. B) Ninhydrin stain.

We submitted the aqueous fraction for high-resolution mass spectrometry analysis and the analysis indicated the correct glycerophoserine product **(2)** size (M-1 258.0384 m/z) (Figure 4.27) however glycerol-3-phosphate **(21)** (M-1 171.00061 m/z and 2M-1 343.0202 m/z) (Figure 4.28) and other unknown by-products were also identified.

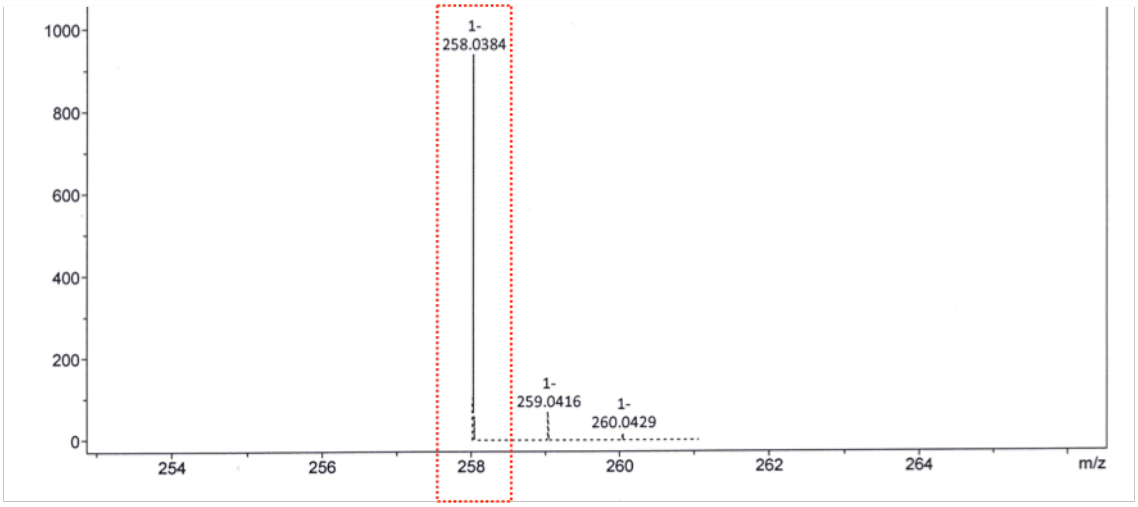


Figure 4.27: Analysis of base hydrolysis products of phosphatidyl-L-serine by high-resolution mass spectrometry. Glycerophoserine (GPS) **(2)** was identified (M-1 258.0384 m/z)

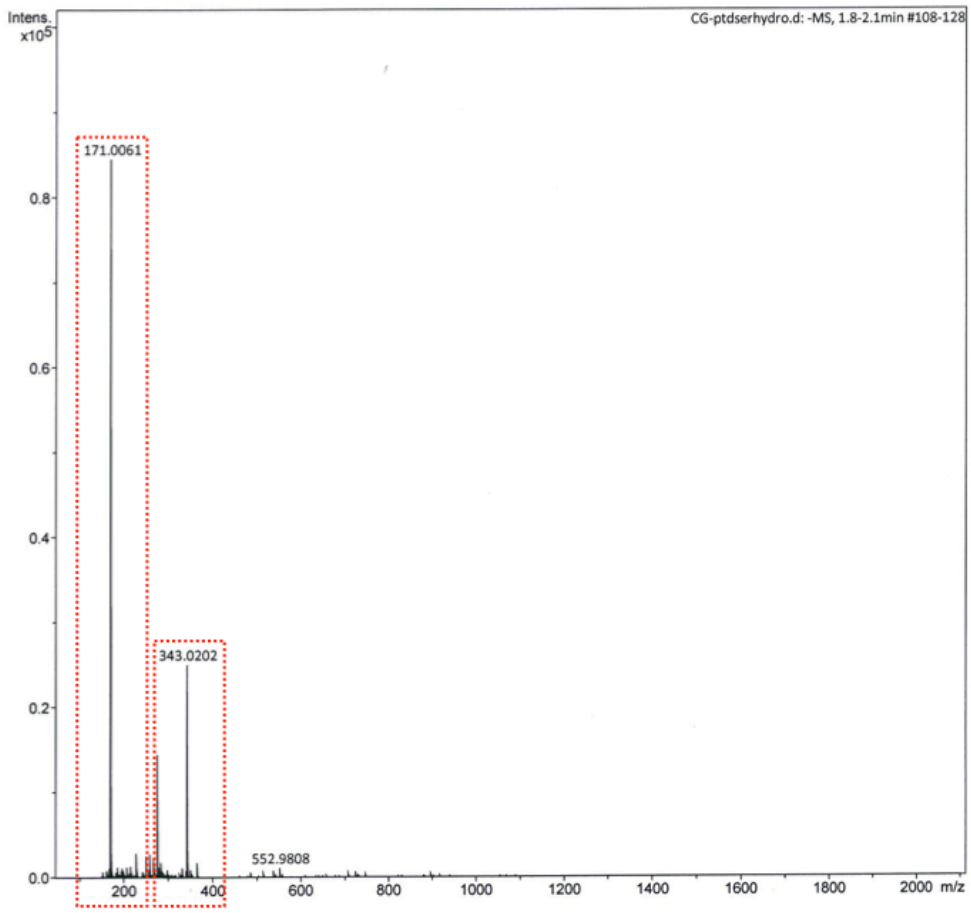


Figure 4.28: Analysis of base hydrolysis products of phosphatidyl-L-serine by high-resolution mass spectrometry.

Hydrolysis of the phosphodiester bond using the base NaOH likely resulted in the formation of by-products **(21)** and **(22)** (Figure 4.29) therefore instead we decided to attempt a more specific enzymatic method.

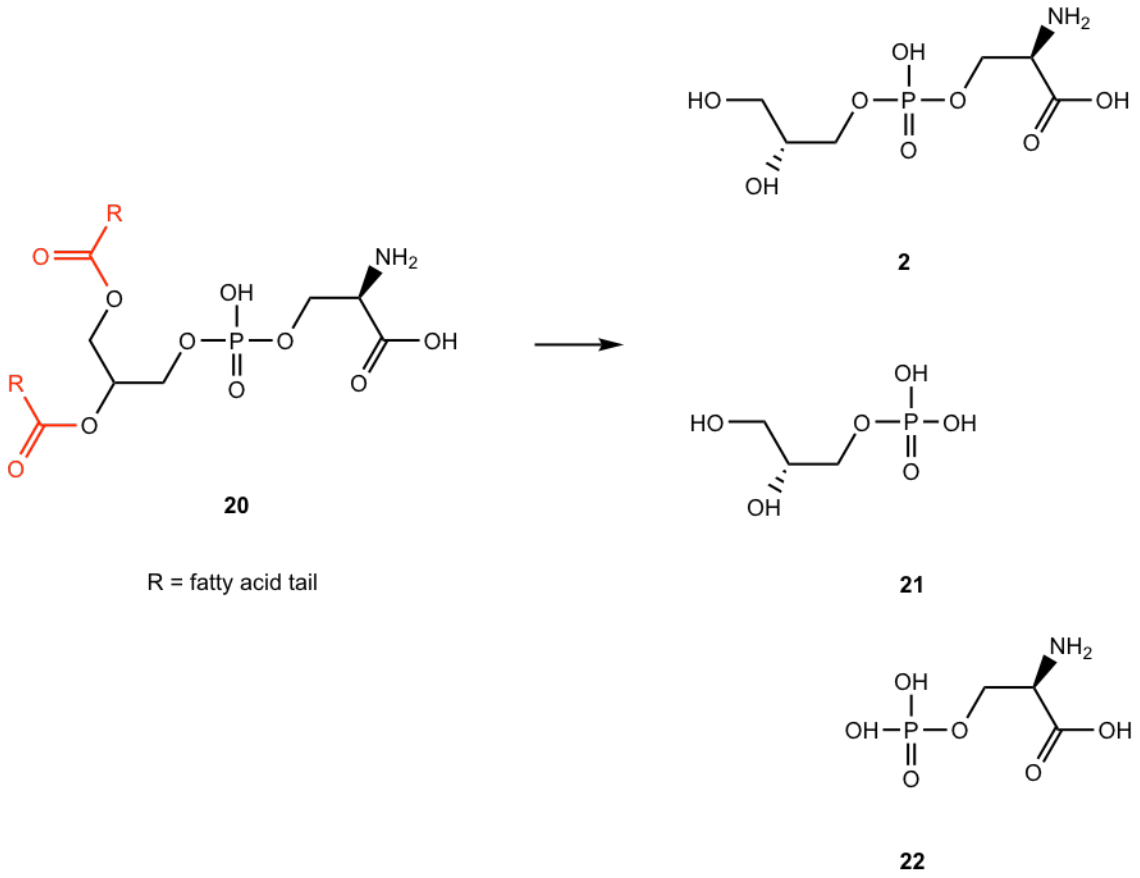


Figure 4.29: By-products produced from base hydrolysis of phosphatidyl-L-serine (20). Glycerophoserine (GPS) **(2)** was identified (M-1 258.0384 m/z) however also glycerol-3-phosphate (G3P **(21)** was identified (M-1 171.00061 m/z and 2M-1 343.0202 m/z). Phosphoserine **(22)** may have been produced however could not be detected.

4.4.3 Enzymatic production of glycerophosphodiesters

As all the previous routes did not result in the correct products we resorted to using an enzymatic method based on a previous study (Bang et al., 2016). The enzyme used was phospholipase A_1 from *Aspergillius oryzae* that specifically cleaves the fatty acid tail at the *sn*-1 position of phospholipids (Chapter 1, 1.1.23). This leaves the fatty acid tail at position *sn*-2 intact. However in the previous study (Bang et al., 2016) it was found that if the phospholipid and enzyme mixture was incubated at 50 °C the fatty acid tail at position *sn*-2 migrates to *sn*-1 thereby enabling both fatty acid tails to be cleaved using the same enzyme (Bang et al., 2016). The enzymatic reaction is described (Figure 4.30). Similar to the base hydrolysis method the fatty acid tails are soluble in chloroform whereas the glycerophosphodiester product is soluble in H_2O therefore no further purification steps would be required.

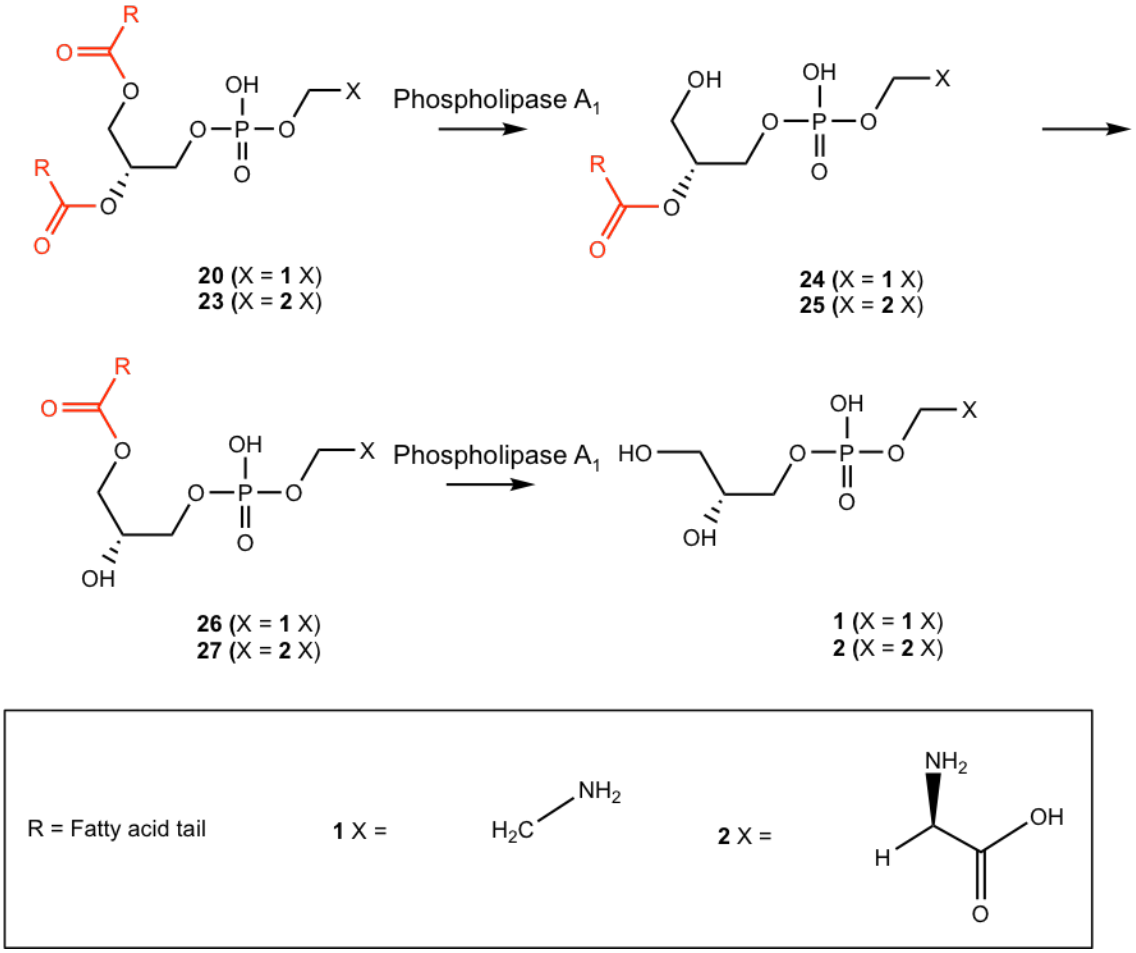


Figure 4.30: Phospholipase A_1 hydrolysis of 1,2-Dipalmitoyl-*sn*-glycero-3-phosphoethanolamine (20) and 1,2-Diacyl-*sn*-glycero-3-phospho-L-serine (23). R refers to the fatty acid tail at position *sn*-1 or *sn*-2 and X refers to the lipid head-group. X refers to the head-group. At 50 °C the fatty acid tail at position *sn*-2 **(24)** and **(25)** migrates to *sn*-1 **(26)** and **(27)**.

The following commercially available phospholipids were digested: 1,2-Dipalmitoyl-*sn*-glycero-3-phosphoethanolamine **(20)** (50 mg) and 1,2-Diacyl-*sn*-glycero-3-phospho-L-serine **(23)** (25 mg) (Figure 4.30). We also digested 2,3-Dipalmitoyl-*sn*-glycero-1-phosphocholine **(28)** (12.5 mg) to produce the unnatural glycerophosphocholine enantiomer (Figure 4.31).

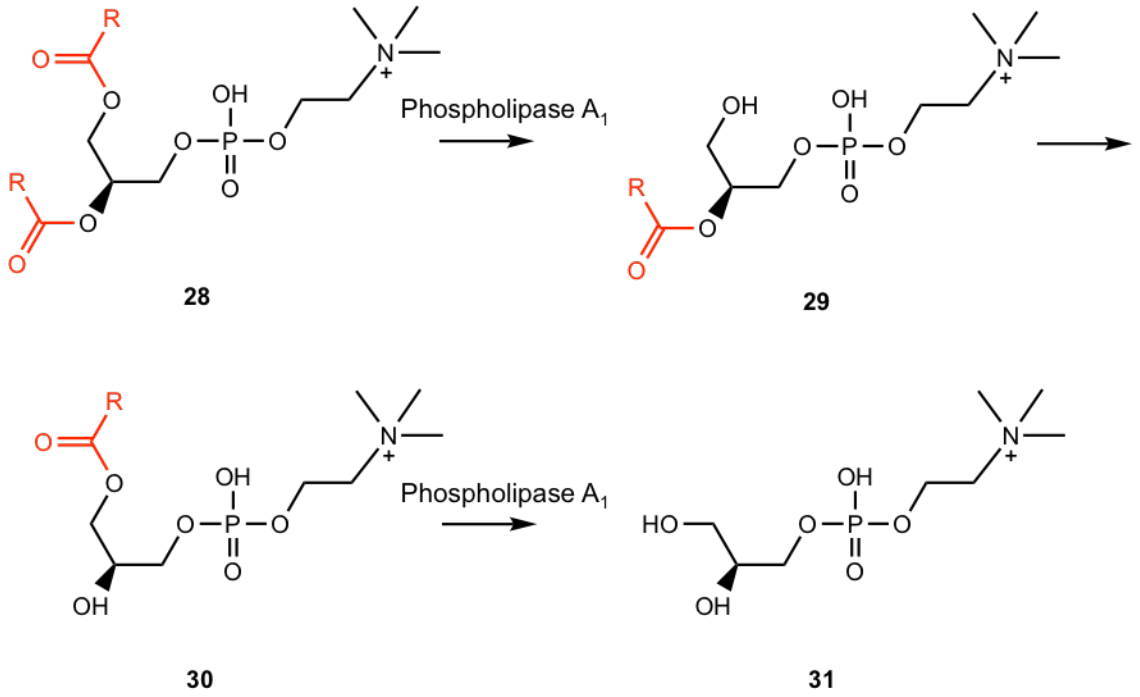


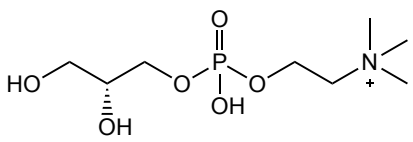
Figure 4.31: Phospholipase A_1 hydrolysis of, 3-Dipalmitoyl-*sn*-glycero-1-phosphocholine (28). R refers to the fatty acid tail at position *sn*-1 or *sn*-2 and X refers to the lipid head-group. At 50 °C the fatty acid tail at position *sn*-2 **(29)** migrates to *sn*-1 **(30)**.

The phospholipids were dissolved in organic-aqueous media (800 μL hexane, 138 μL H_2O: 5.8:1 ratio) and heated at 50 °C for 10 minutes prior to the addition of Phospholipase A_1 from *Aspergillus oryzae* (1 μL Phospholipase A_1 per 1 mg of phospholipid). The reaction mixtures were heated at 50°C and stirred at 300 rpm for 48 hours. Following incubation the solvent was removed *in vacuo*, the reaction mixtures were dissolved in water (5 mL) and the reaction mixtures were extracted with chloroform (3 x 25 mL) to remove the chloroform soluble fatty acid tails from the H_2O soluble glycerophosphodiesters. The aqueous fractions containing the glycerophosphodiesters and phospholipase A_1 enzyme were passed over a centrifugal filter unit (Amicon, 10 kDa molecular weight cut off) to remove phospholipase A_1. The products were concentrated and yielded a colourless oil: 9.4 mg glycerophosphoethanolamine **(1)**, yellow oil: 3 mg glycerophosphoserine **(2)** and a colourless oil 1.5 mg glycerophosphocholine enantiomer **(31).** The products were characterised by proton, carbon and phosphate NMR (Appendix 4.7) (Chapter 2, 2.7.5). By NMR glycerophosphoethanolamine **(1)** was 90 % pure, glycerophosphoserine **(2)** was 80 % pure and the glycerophosphocholine enantiomer **(31)** was 33 % pure. The only impurities appeared to be glycerol.

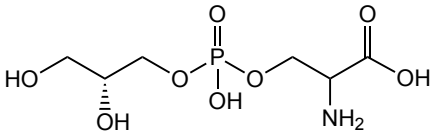
Now that we had all of the ligands we were ready to test binding of the enzymatically produced alternative glycerophosphodiesters to *Mtb* UgpB by MST (4.5).

4.5 Alternative glycerophosphodiester binding

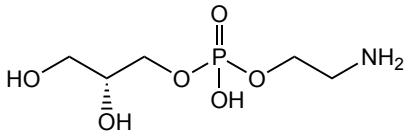
4.5.1 Binding affinity analysis of *Mtb* UgpB with alternative glycerophosphodiesters by microscale thermophoresis



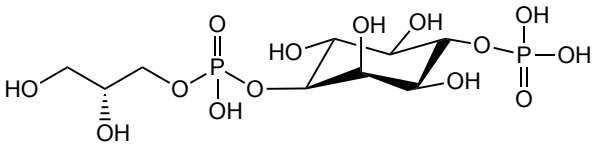
Glycerol-3-phosphocholine (GPC)



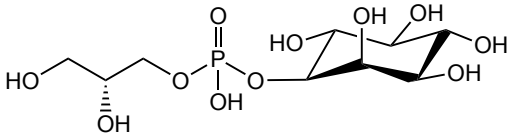
Glycerophosphoserine (GPS)



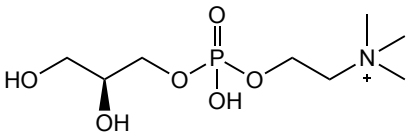
Glycerophosphoethanolamine (GPE)



Glycerophosphoinositol-4-phosphate (GPI(4)P)



Glycerophosphoinositol (GPI)



Glycerophosphocholine enantiomer

Figure 4.32: Structures of the alternative glycerophosphodiesters tested for binding to *Mtb* UgpB by microscale thermophoresis. Glycerophosphoserine (GPS), glycerophosphoethanolamine (GPE) and the glycerophosphocholine enantiomer were enzymatically produced. Glycerol-3-phosphocholine (GPC), glycerophosphoinositol-4-phosphate (GPI(4)P) and glycerophosphoinositol (GPI) were commercially available.

The glycerophosphodiester compounds that were produced enzymatically were tested for the ability to bind to *Mtb* UgpB by MST (Figure 4.32): GPE (90 % pure), GPS (80 % pure) and GPC enantiomer (33 % pure). *Mtb* UgpB was prepared for MST analyses as described previously (4.3.12) (Chapter 2, 2.5.12). Both of the alternative glycerophosphodiesters were found to bind and the highest affinity was for GPS with a K_d of 14.9 μM (Figure 4.33A) (Table 4.3) and a slightly lower binding affinity for GPE with a K_d of 74.7 μM (Figure 4.33B) (Table 4.3). Binding of *Mtb* UgpB to the commercially available glycerophosphodiesters: GPI(4)P and GPI was also investigated by MST. These compounds were selected because the lipid strips indicated that *Mtb* UgpB binds to PtdIns(4)P (4.3.8). GPI(4)P had a K_d of 289.8 μM (Figure 4.33C) (Table 4.3) and was slightly higher compared to the binding affinity of GPI with a K_d of 1053.2 μM (Figure 4.33D) (Table 4.3). Together these results indicated that *Mtb* UgpB is able to bind to alternative glycerophosphodiesters and the

highest binding affinities were for GPS and GPE. The result of *Mtb* UgpB binding to GPI(4)P with a higher affinity compared to GPI also indicated that the larger phosphorylated inositol head-group was preferred.

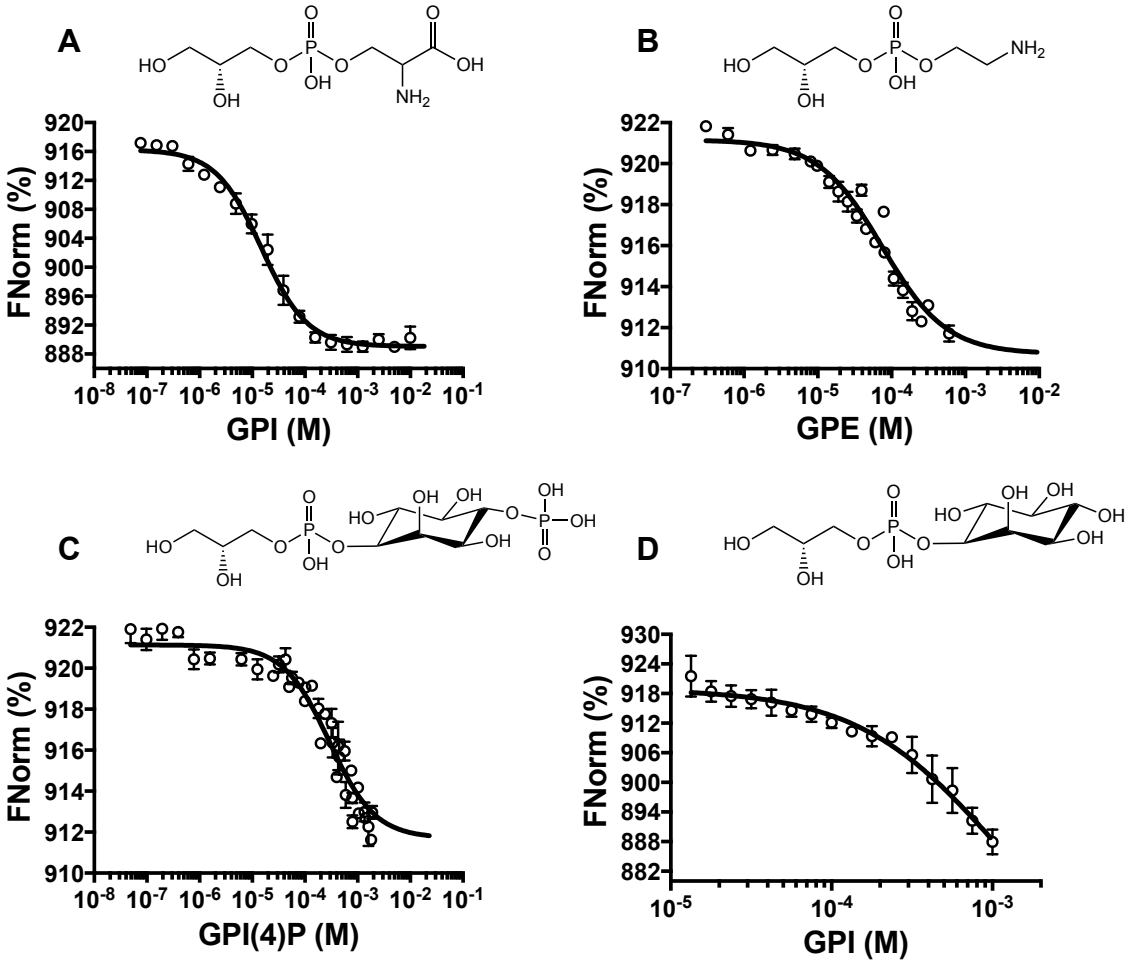


Figure 4.33: Microscale thermophoresis of *Mtb* UgpB and alternative glycerophosphodiesters. FNorm (%) is the normalised fluorescence signal of the change in MST signal. Error bars represent standard deviations from at least three independent experiments. A) Glycerophosphoserine (GPS). B) Glycerophosphoethanolamine (GPE). C) Glycerophosphoinositol-4-phosphate (GPI(4)P). D) Glycerophosphoinositol (GPI).

We also wanted to investigate the importance of the stereo-configuration of the glycerol group of GPC by testing the unnatural enantiomer of GPC. As based on the structure of GPC bound *Mtb* UgpB (Chapter 5, 5.3.6), the 2-hydroxyl of the glycerol group interacts with Asp102 in the binding pocket. Therefore if the orientation of the 2-hydroxyl of the glycerol group is altered it may exhibit an altered binding affinity. The structures of the unnatural enantiomer of GPC and natural GPC are shown (Figure 4.34).

Glycerophosphocholine enantiomer **Glycerol-3-phosphocholine (GPC)**

Figure 4.34: Structure of the unnatural glycerophosphocholine (left) enantiomer compared to GPC (right).

The GPC enantiomer was found to bind to *Mtb* UgpB by MST with a K_d of 1.1 mM (Figure 4.35), approximately 300-fold lower compared to the standard GPC K_d of 3.6 μM. This result is likely to reflect the importance of the orientation of the 2- hydroxyl of the GPC glycerol group for binding to Asp102 in the *Mtb* UgpB binding pocket. This is further supported because when Asp102 is mutated to Ala, binding to the standard GPC is also abolished (4.3.13).

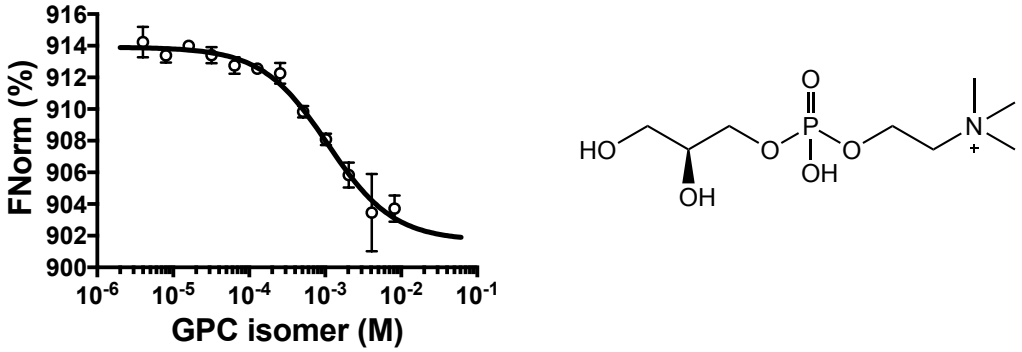


Figure 4.35: Microscale thermophoresis *Mtb* UgpB and glycerophosphocholine enantiomer. FNorm (%) is the normalised fluorescence signal of the change in MST signal. Error bars represent standard deviations from at least three independent experiments.

A summary of the *Mtb* UgpB MST binding affinity results is shown (Table 4.3). Overall GPC had the highest affinity with a K_d of 3.6 μM. This was closely followed by GPS that had a K_d of 14.9 μM and suggests that the serine head-group is well tolerated. GPE had a slightly weaker binding affinity with a K_d of 74.7 μM, approximately 20-fold lower compared to GPC. This is potentially because the ethanolamine head-group is smaller and *Mtb* UgpB appears to have a preference for larger head-groups. GPI(4)P also bound with a K_d of 289.8 μM, approximately 80-fold lower compared to GPC. GPI(4)P binding is also supported by STD-NMR data (Chapter 5, 5.3.13) that interestingly revealed that regions of the inositol phosphate head-group are involved in binding recognition. In comparison, GPI, which is smaller and lacks a phosphorylated inositol group had a weaker affinity with a K_d of 1053.2 μM, approximately 290-fold lower compared to GPC. However the weakest affinity was for the GPC enantiomer with a K_d of 1080 μM and is potentially due to the different stereo configuration of the 2-hydroxyl

glycerol group. Of the mutants tested, Tyr78Ala and Asp102Ala did not bind to either GPS or GPE and this result was identical to the result found previously (4.3.13) for GPC, and clearly suggests that Tyr78 and Asp102 are important for glycerophosphodiester binding recognition. However, Ser153Ala was found to bind to GPS with a K_d of 102.5 μM approximately 7-fold lower compared to apo-*Mtb* UgpB. Ser153Ala also had a weaker affinity for GPC with a K_d of 309.8 μM (4.3.13), yet surprisingly Ser153Ala did not bind to GPE. This was potentially due to the smaller ethanolamine head-group.

Table 4.3: Summary of all *Mtb* UgpB MST binding affinity data.

Protein	Substrate	K_d (μM)
Mtb UgpB	GPC	3.6 ± 0.5
Mtb UgpB	GPC*	5.2 ± 1.6
Mtb UgpB	GPC enantiomer	1080 ± 188.3
Mtb UgpB	GPS	14.9 ± 1.6
Mtb UgpB	GPE	74.7 ± 13.9
Mtb UgpB	GPI(4)P	289.8 ± 54.1
Mtb UgpB	GPI	1053.2 ± 313.4
Mtb UgpB Y78A	GPS	-
Mtb UgpB Y78A	GPE	-
Mtb UgpB D102A	GPS	-
Mtb UgpB D102A	GPE	-
Mtb UgpB S153A	GPS	102.5 ± 16.4
Mtb UgpB S153A	GPE	-

(-) = no binding detected, ± SD from three independent experiments. GPC: glycerol-3-phosphocholine, GPS: glycerophosphoserine, GPE: glycerophosphoethanolamine, GPI(4)P: glycerolphosphoinositol-4-phosphate, GPI: glycerophosphoinositol. * refers to the GPC competition experiment.

In conclusion, *Mtb* UgpB has demonstrated broad substrate selectivity and was found to be promiscuous for binding to a wide range of glycerophosphodiester substrates.

4.5.2 Microscale thermophoresis competition experiment

To investigate if the presence of G3P altered GPC binding a competition experiment was carried out (Figure 4.36). As it wasn't clear why GPC bound to *Mtb* UgpB however G3P did not bind to *Mtb* UgpB. Even though both compounds have a glycerol and phosphate moiety and glycerol is the main recognition element. In this experiment 100 μM of G3P was included in each assay and the concentration of GPC was varied from 0.1-11 μM. The binding affinity for GPC in the presence of G3P was 5.2 μM (Figure

4.36) (Table 4.3) and was comparable to GPC only of 3.6 μM (4.3.12). This result further suggested that GPC binds preferentially to *Mtb* UgpB as the presence of G3P did not alter the binding affinity to GPC. This experiment also supports previous thermal shift assay data (4.3.4) that indicated GPC binds however G3P does not bind to *Mtb* UgpB.

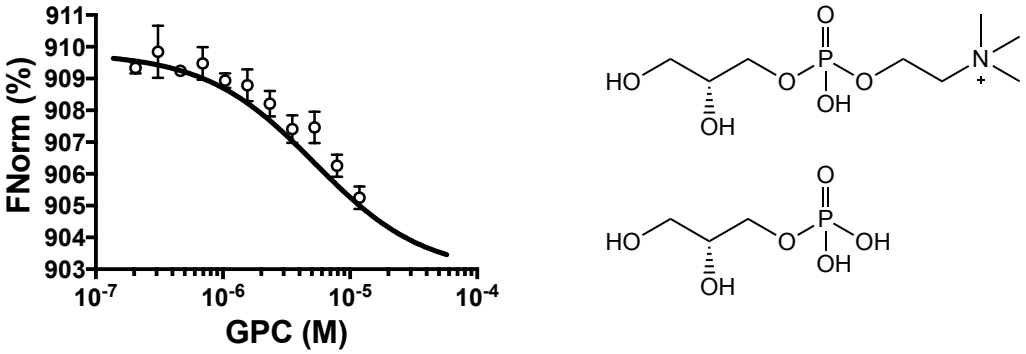


Figure 4.36: Microscale thermophoresis *Mtb* UgpB and 100 μM glycerol-3-phosphate and glycerophosphocholine. FNorm (%) is the normalised fluorescence signal of the change in MST signal. Error bars represent standard deviations from at least three independent experiments.

4.6 Discussion

A thermal shift assay testing a panel of carbohydrates, amino acids, glycerophosphodiester variations and antibiotics found GPC to have the highest thermal shift (+7.3 °C) (4.3.4) Binding to GPC was determined by MST to have a K_d of 3.6 μM (4.3.12) and was comparable to other substrate binding proteins of the ATP-binding cassette family (Berntsson et al., 2010). The MST result was also similar to a previous study that measured binding of *Mtb* UgpB to GPC by isothermal calorimetry (ITC) and found GPC to have a K_d of 27.3 μM (Jiang et al., 2014). By MST, binding to G3P or phosphocholine could also not be detected and is in good agreement with the previous study (Jiang et al., 2014). Together, our MST data suggested that the combined glycerol, phosphate and choline moieties were required for binding recognition.

The GPC bound *Mtb* UgpB X-ray crystal structure was later determined (Chapter 5, 5.3.6) and revealed 8 binding site residues that interacted with GPC. Based on our structure, the 8 binding site residues of *Mtb* UgpB were individually mutated to alanine by site directed mutagenesis and mutant derivatives of *Mtb* UgpB were successfully purified (Chapter 3, 3.3.10). MST analysis found that all of the mutant derivatives of *Mtb* UgpB either did not bind to GPC or had a lower binding affinity compared to apo-*Mtb* UgpB. Leu205Ala and Ser153Ala had reduced binding affinities with a K_d of 161.7 μM and a K_d of 309.8 μM respectively (4.3.13). This result contrasted with the previous study (Jiang et al., 2014) that did not identify binding of Leu205Ala to GPC.

Interestingly, based on the GPC bound *Mtb* UgpB structure (Chapter 5, 5.3.6), the choline head-group of GPC did not appear to interact with *Mtb* UgpB. We therefore hypothesised that alternative glycerophosphodiester head-groups could also bind. Glycerophosphodiesters could potentially be derived from phospholipids of the macrophage membrane or lung surfactant (Sahu et al., 1977) (Fessler et al., 2016). We initially investigated binding to lipid strips containing phospholipids with different head-groups. The results of the lipid strips found that *Mtb* UgpB had a preference for phosphoinositides, phosphatidylserine (4.3.7) and sulfatide (4.3.9). However, surprisingly binding to phosphatidylcholine or phosphatidylinositol was not found. In hindsight the lipid strips may have not been the best assay to use because the fatty acid tail was ester bonded to the glycerol group and therefore blocked access to the glycerol moiety.

Instead, we decided to directly investigate the interactions of alternative glycerophosphodiesters to *Mtb* UgpB. Therefore we tested GPI, GPI(4)P, GPE and GPS. We tested GPI because GPI is the head-group of phosphatidylinositol and phosphatidylinositol mannosides (PIMs), which are found in abundance in the *Mycobacteria* inner membrane (Bansal-Mutalik et al., 2014). We also tested GPE because phosphatidylethanolamine is also found in the *Mycobacteria* inner membrane (Bansal-Mutalik et al., 2014). All of the lipids of which these head-groups are derived are also found in macrophage membranes (Sahu et al., 1977) and therefore could be produced by the activities of *Mtb* or host derived phospholipase A_1 and A_2 enzymes (Brust et al., 2011) (Schue et al., 2010) (B. E. Schneider et al., 2014).

GPI and GPI(4)P were commercially available however we had to produce GPS and GPE. To produce GPS and GPE we decided to use a synthetic route based on previous studies (R.-R. Tang et al., 2005) (Gagnon et al., 2017). In order to direct the reaction to produce the glycerophosphodiesters we had to use protected reagents: (S)-isopropylideneglycerol, phenyldichlorophosphate, *N*-Boc-ethanolamine and *N*-Boc-L-serine methyl ester. We optimised the reaction by altering the base and the order of reagent additions however in all cases we produced several by-products that could not be effectively separated by column chromatography. Therefore we decided instead to start with the intact phospholipid and use a chemical or enzymatic method to cleave the fatty acid tails leaving the glycerophosphodiester intact. The base hydrolysis method (Kopp et al., 2010) produced the correct glycerophosphodiester products however it also produced several by-products therefore we decided to use a more specific enzymatic route (Bang et al., 2016). The glycerophosphodiesters, GPS and GPE were successfully produced enzymatically and we also produced the unnatural enantiomer of GPC by enzymatically digesting 2,3-Dipalmitoyl-sn-glycero-1-phosphocholine.

The alternative glycerophosphodiesters were also found to bind by MST (4.5.1). GPS was found to bind with a K_d of 14.9 μM, similar to GPC and GPE was found to bind with a K_d of 74.7 μM approximately 5-fold lower compared to GPC. The GPC enantiomer indicated a considerably lower binding affinity compared to the standard GPC, with a K_d of 1.08 mM and suggested that the stereo-configuration of the 2-hydroxyl of the glycerol moiety was important for binding recognition. GPI(4)P bound with a K_d of 289.8 μM, approximately 80-fold less compared to GPC and GPI bound with a K_d of 1053 μM. These results clearly suggested that *Mtb* UgpB is able to bind and recognise a broad range of glycerophosphodiester substrates containing different head-groups. The order of binding preference is: GPC>GPS>GPE>GPI(4)P>GPI>GPC enantiomer.

Interestingly by metabolic profiling, GPC is increased in *Mtb* infected guinea pig granuloma (Somashekar et al., 2011) therefore this places the *Mtb* bacilli in close proximity to a potential nutrient source and substrate of *Mtb* UgpB. However we have also found that *Mtb* UgpB may also be adapted to bind and recognise alternative glycerophosphodiesters that are potentially derived from macrophage membranes in order to recycle nutrients from this scarce environment.

4.7 Appendix

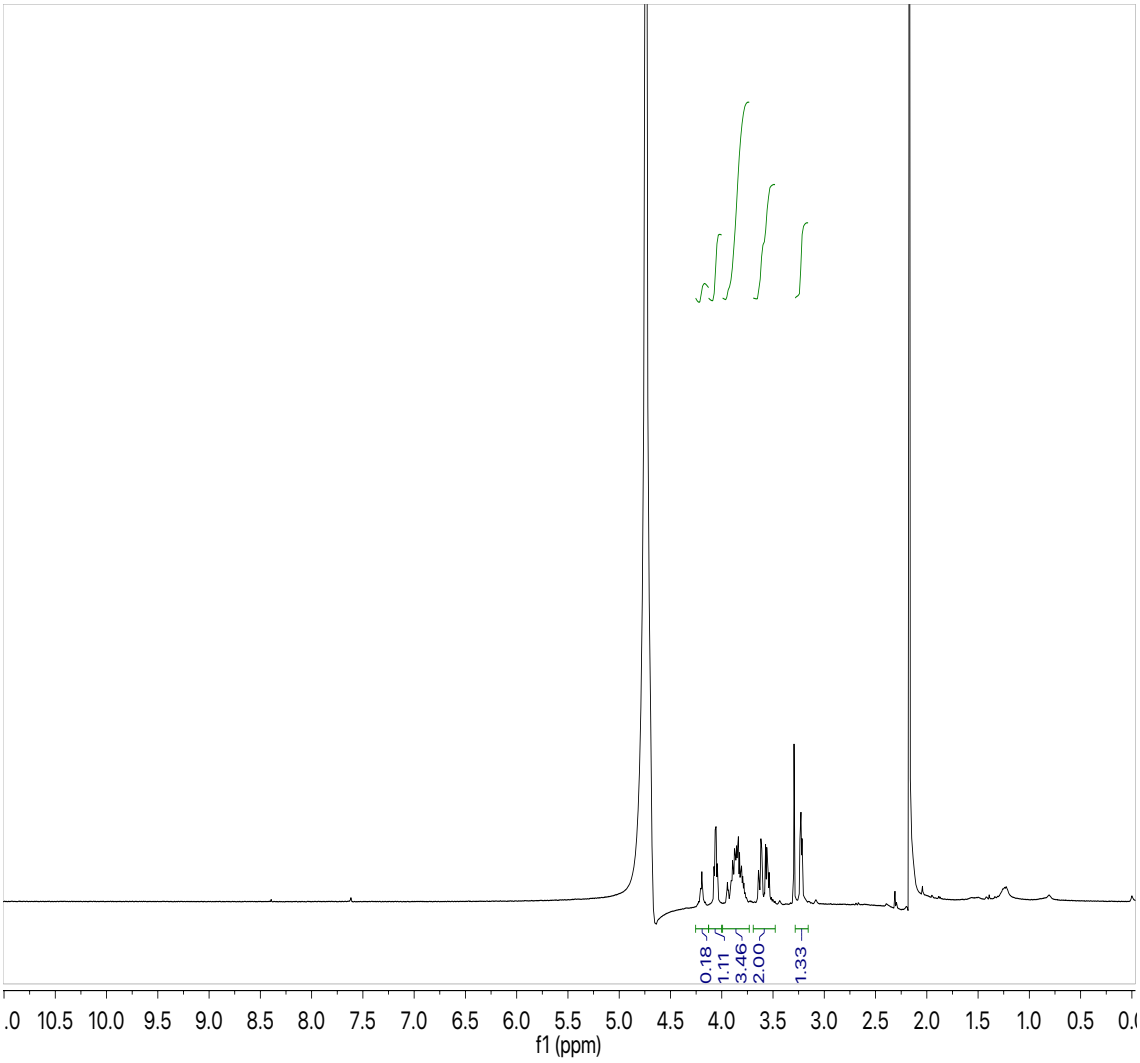


Figure 4.37: GPS ^{1}H NMR. (400MHz, D_2O) δ_{ppm} 4.17 – 4.26 (2H, m, POC$\underline{H}_2$CHN), 4.01 – 4.08 (1H, m, POCH$_2$C$\underline{H}$N), 3.71 – 3.93 (3H, m, 3H, m, POC$\underline{H}_2$C$\underline{H}$CH$_2$), 3.48-3.65 (2H, m, POCH$_2$CHC$\underline{H}_2$).

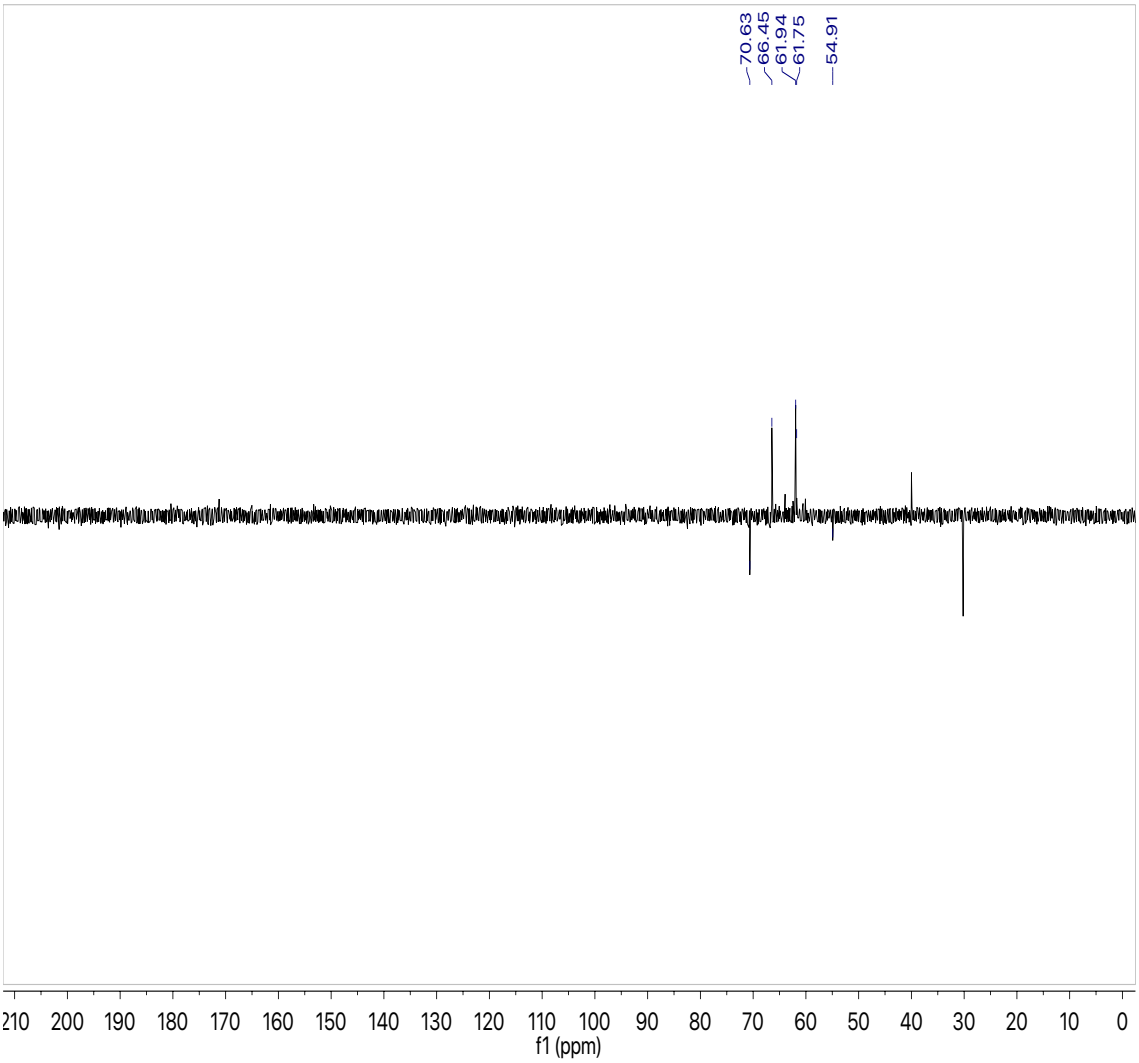


Figure 4.38: GPS ^{13}C NMR. (100MHz, D_2O) δ_{ppm} 178.5 ($\underline{C}$=O), 70.6 ($CH_2\underline{C}H(OH)CH_2$), 66.5 ($O\underline{C}H_2$), 63.7 ($PO\underline{C}H_2CHN$), 62.0 ($O\underline{C}H_2$), 54.4 ($POCH_2\underline{C}HN$).

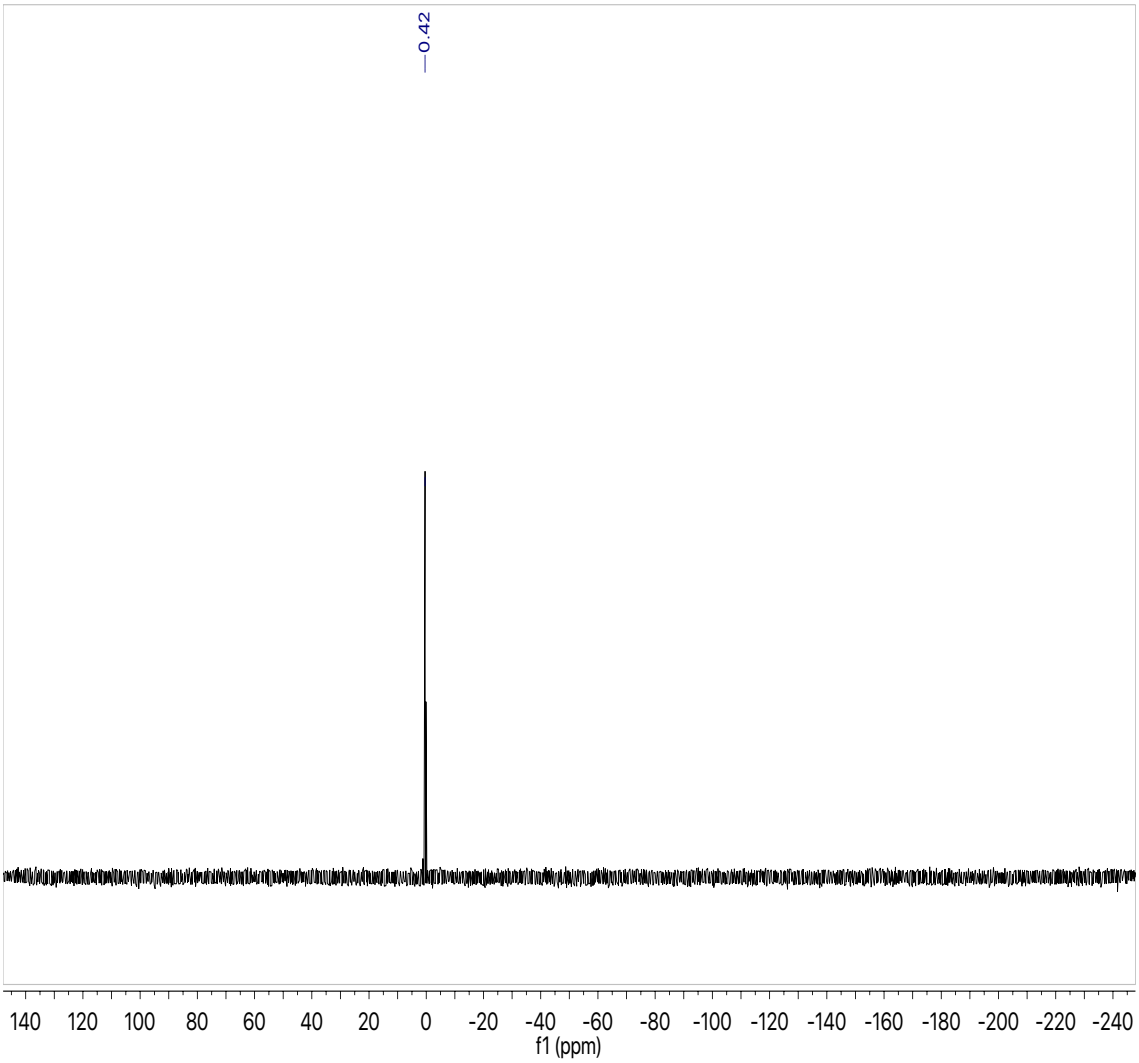


Figure 4.39: GPS ^{31}P NMR. (161 MHz, D_2O) δ_{ppm} 0.08.

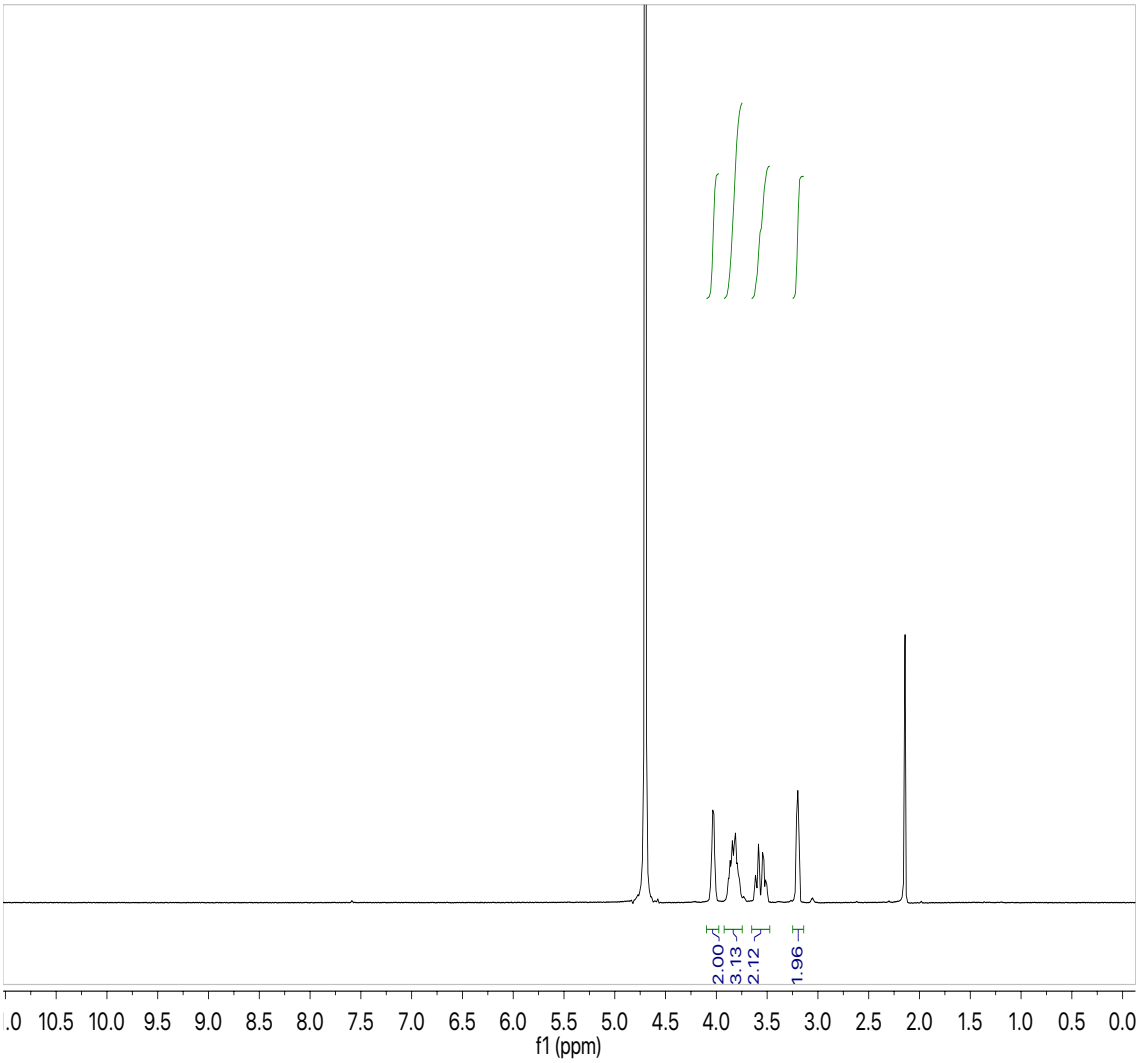


Figure 4.40: GPE: ^{1}H NMR. (400MHz, D_2O) δ_{ppm} 3.99 – 4.08 (2H, m, POC<u>H</u>$_2$C<u>H</u>$_2$N), 3.75- 3.91 (3H, m, POC<u>H</u>$_2$C<u>H</u>C<u>H</u>$_2$), 3.49 – 3.63 (2H, m, POCH$_2$CHC<u>H</u>$_2$), 3.20 (2H, t, *J* = 5.0 Hz, POCH$_2$C<u>H</u>$_2$N).

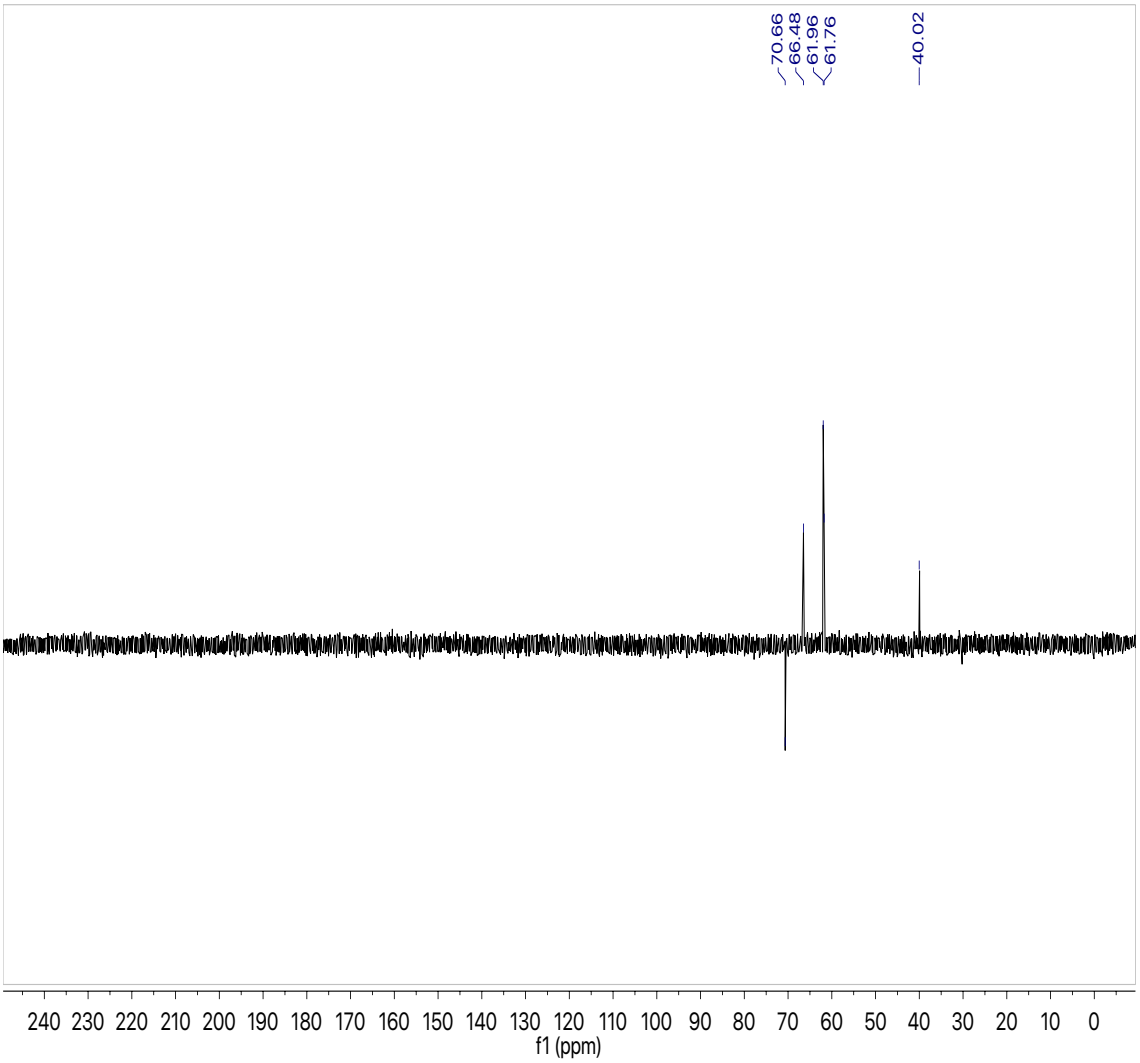


Figure 4.41: GPE ^{13}C NMR. (100MHz, D_2O) δ_{ppm} 70.7 (C̲H), 66.5 (OC̲H_2), 62.0 (OC̲H_2), 61.8 (OC̲H_2), 40.0 (NC̲H_2). ^{31}P NMR (161 MHz, D_2O) δ_{ppm} 0.42.

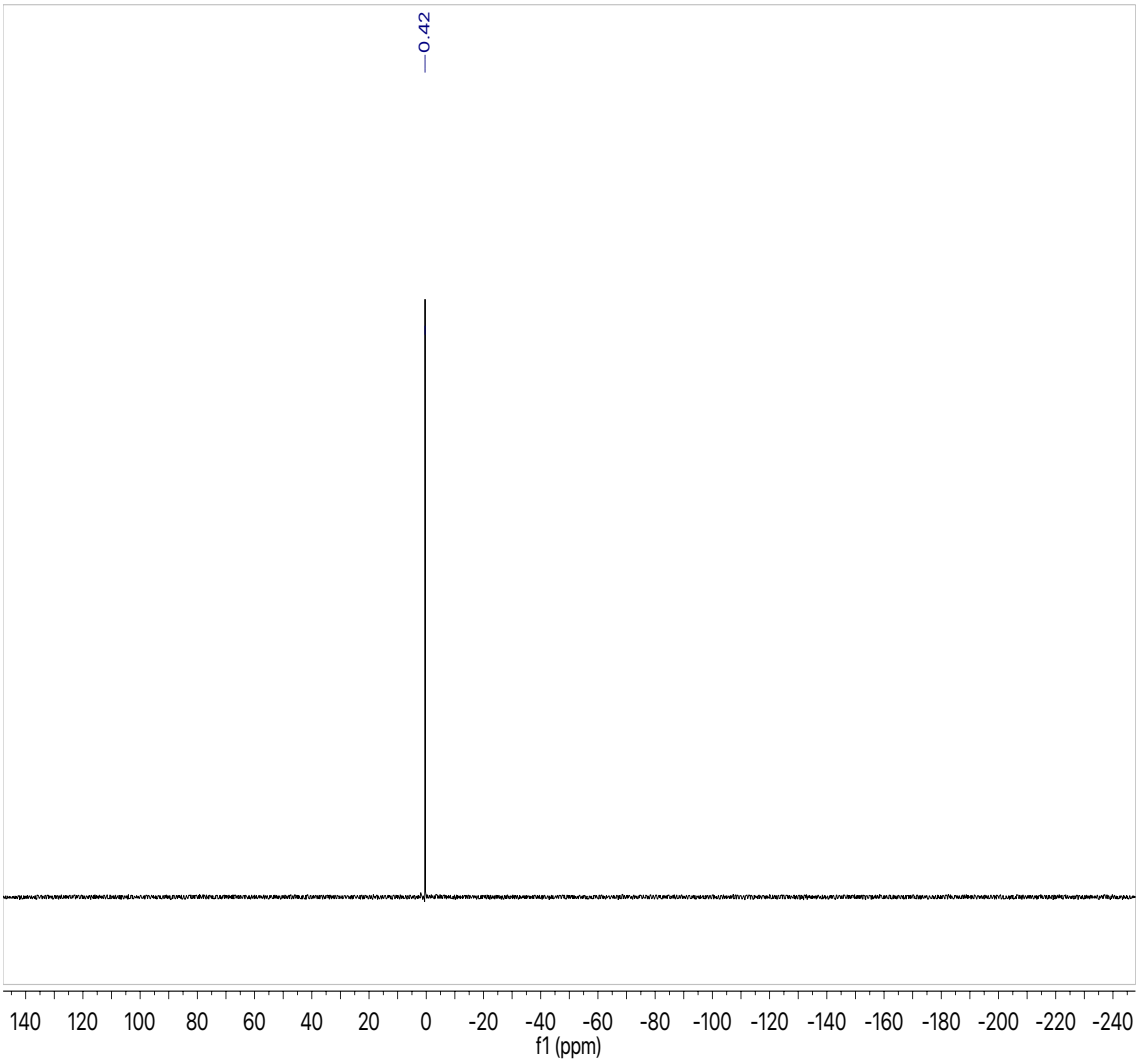


Figure 4.42: GPE ^{31}P NMR. (161 MHz, D_2O) δ_{ppm} 0.42.

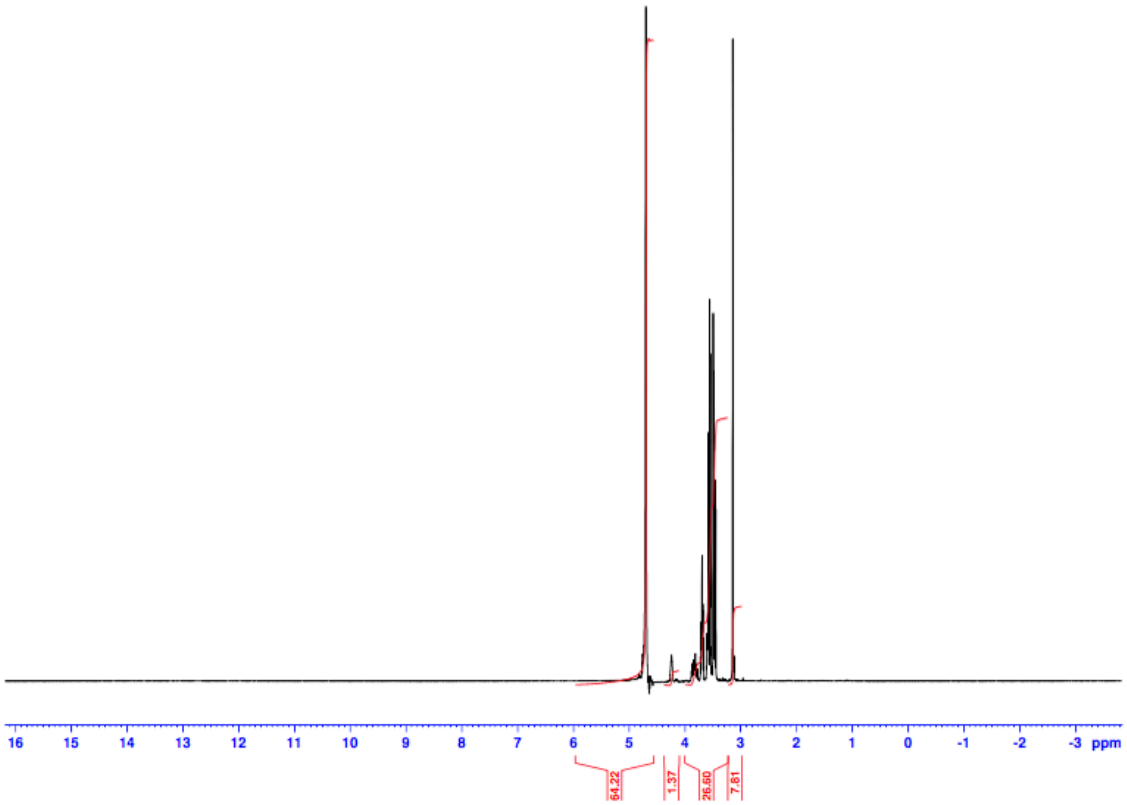


Figure 4.43: GPC enantiomer: ^{1}H NMR. (400MHz, D_2O) δ_{ppm} 4.18 – 4.29 (2H, m, $POC\underline{H}_2CH_2N$), 3.74 – 3.90 (3H, m, $POC\underline{H}_2C\underline{H}CH_2$), 3.46 – 3.63 (4H, m, $POCH_2CHC\underline{H}_2$ + $POCH_2C\underline{H}_2N$), 3.14 (9H, s, $C\underline{H}_3$).

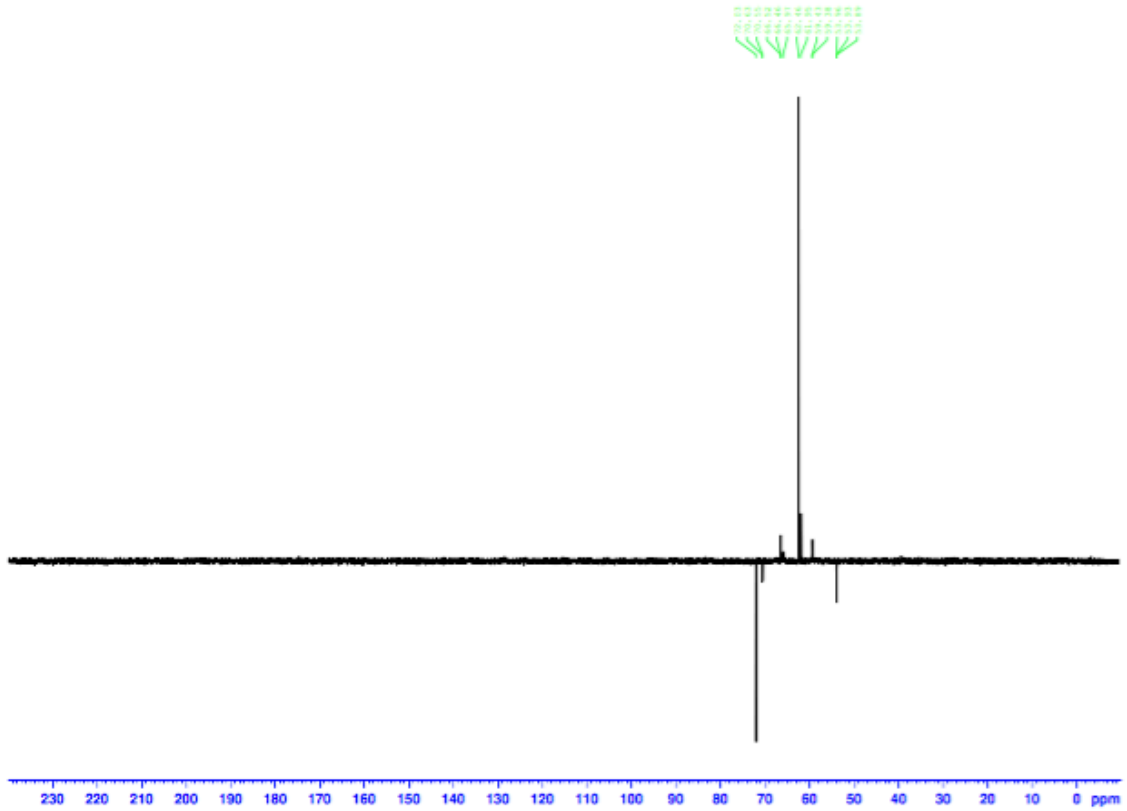


Figure 4.44: GPC enantiomer ^{13}C NMR. (100MHz, D_2O) δ_{ppm} 70.6 ($CH_2\underline{C}H(OH)CH_2$), 66.5 ($O\underline{C}H_2$), 66.0 ($O\underline{C}H_2$), 62.0 ($O\underline{C}H_2$), 59.4 ($N\underline{C}H_2$), 53.9 ($\underline{C}H_3$).

Chapter 5: Structural studies of *Mycobacterium tuberculosis* UgpB

5.1 Introduction

5.1.1 Introduction

Structural determination of proteins is a powerful method to elucidate their functions. It reveals details about the secondary, tertiary and quaternary structures, information about how the protein folds, alters conformational state and can identify regions that interact with other macromolecules and small molecule ligands. To date, 2440 *Mycobacterium tuberculosis* (*Mtb*) protein structures are available (RCSB, 2019) that have been solved by four different methods. X-ray diffraction is the most widely used method and is responsible for 2365 structures whereas cryo-electron microscopy, solid-state NMR and solution NMR, combined are responsible for only 75 structures/partial structures. Investigating how small molecule ligands target, alter the activity or inhibit proteins is particularly useful because they are widely used as drugs and antibiotics to treat bacterial infections. Solving the structures of ligand bound targets is therefore a powerful method to identify new drug targets and investigate how to modify and improve existing small molecules. It also provides a functional insight to understand the underlying biochemistry.

Various methods are available to study the structures of molecules. X-ray crystallography is capable of producing the highest atomic resolution structures however obtaining the crystals and solving the structures can be challenging. With the recent development of new algorithms and direct electron detectors, cryo-electron microscopy is now approaching the resolution levels of X-ray diffraction (Lyumkis, 2019) however the highest resolution structures are still limited to complexes or large molecules ≥ 200 kDa. Solid-state NMR has been developed to determine the structure and conformational dynamics of membrane proteins within the native membrane environment (X. Zhao, 2012). For proteins that are difficult to crystallise, solution NMR methods are also available, however the size limit is approximately 30 kDa (Billeter et al., 2008). More recently, saturation transfer difference NMR (STD-NMR) has been developed to elucidate protein-glycan interactions (Blaum et al., 2018) and differential epitope mapping (DEEP) STD-NMR (Monaco et al., 2017) has been developed to reveal information about how protein amino acid residues interact with particular ligand moieties.

5.1.2 Structures of ABC transporter SBP's

Substrate binding proteins (SBP's) of ATP-binding cassette transporters (ABC) vary in size from around 25-70 kDa (Quiocho et al., 1996). All are made of two globular α/β domains connected by flexible hinge regions that enable them to exist in at least four possible conformational states, open un-liganded, open liganded, closed un-liganded and closed liganded (C. Tang et al., 2007). SBP's are able to bind to ligands *via* a flexible hinge resulting in a conformational closure of the two α/β domains around the ligand by a mechanism termed the Venus Fly-trap (Mao et al., 1982). SBP's were previously classified as either class I, II or III due to sequence similarity and core β sheet topology in the two α/β domains (Fukami-Kobayashi et al., 1999). Generally in class I SBP's there are 3 connecting strands whereas there are only 2 connecting strands in class II SBP's. Class III SBP's are different as they instead have a single α-helix linking the two-α/β domains. However based on the huge increase in available SBP structures they were reclassified based on structural alignment to six different clusters A-F (Figure 5.1) (Berntsson et al., 2010). Cluster A (class III) have a single α-helix functioning as the hinge and they are part of ABC importers. Cluster B (class I) have three interconnecting strands between the two-α/β domains. Cluster C (class II) have an additional domain between the two-α/β domains. Cluster D (class II) have a very short hinge region consisting of two strands 4-5 amino acids long. Cluster E (class II) associate with TRAP or TT transporters that use electrochemical ion gradients for transport. Cluster F (class II) have a longer 8-10 amino acid hinge region containing two segments connecting domains I and domains II.

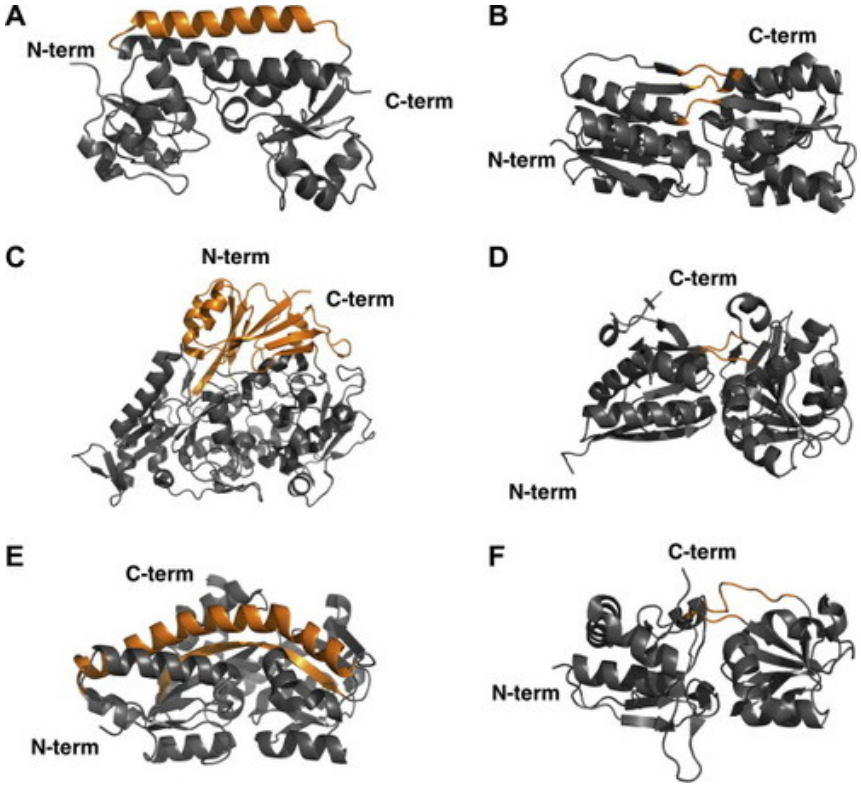


Figure 5.1: Structural classification of substrate binding proteins. Cluster A-F, the distinct structural feature is labelled orange. Clusters A–F were BtuF (PDB 1N2Z), RBP (PDB 1DRJ), OppA (PDB 3DRF), ModA (PDB 1ONR), UehA (PDB 3FXB) and HisJ (PDB 1HSL), respectively (Berntsson et al., 2010). The substrate-binding protein structures differ by the number of interconnecting segments, the length of the flexible hinges and the number of domains.

To date, all SBP's have a highly similar three-dimensional fold despite differences at the amino acid sequence level and this makes it very difficult to infer function based solely on sequence homology. The first SBP crystal structure to be solved was the L-arabinose binding protein of *E. coli* (PDB ID 1ABE) (Quiocho et al., 1974) however now 501 structures have been solved that range from different bacterial species and bind to a wide variety of ligands including carbohydrates, metals, vitamins, amino acids, peptides and alternative metabolites (Scheepers et al., 2016).

5.1.3 Structures of *Mtb* ABC transporter SBP's

To date of the predicted 37 ABC transporters in the *Mtb* genome, including 16 importers and 21 exporters (Braibant et al., 2000), only five *Mtb* SBP structures are available: UspC (Rv2318) (PDB ID 5K2X/5K2), GlnH (Rv0411c) bound to aspartate, asparagine and glutamate (PDB ID 6H1U, 6H20, 6H2T) and UgpB (Rv2833c) (PDB ID 4MFI). Rv2318 encodes *Mtb* UspC the SBP of the *Mtb* UspABC transporter. The structure of *Mtb* UspC with no ligand bound was solved by X-ray diffraction to 1.5 Å resolution and it has the overall architecture characteristic of SBP's with two-α/β domains (Fullam et al., 2016). Although the ligand bound structure could not be

determined a ligand-binding pocket between the two domains was identified to consist of the aromatic side chains of: Trp46, Tyr77, Phe81, Tyr103, Tyr292 and Phe402 that potentially form π-stacking interactions with carbohydrate ligands. High thermal shifts and STD-NMR signals were also identified for D-glucosamine-6-phosphate and chitobiose indicating that *Mtb* UspC had a preference for amino-sugars leading the authors to propose it may be involved in peptidoglycan cell wall fragment recycling (Fullam et al., 2016).

The structure of *Mtb* GlnH in complex with the ligands aspartate, asparagine and glutamate was determined by X-ray crystallography to 1.4-1.7 Å resolution. All structures of *Mtb* GlnH revealed a similar binding mode to the different ligands (Bhattacharyya et al., 2018). The overall architecture of *Mtb* GlnH was similar to other SBP's as the two-α/β domains enclosed the aspartate ligand in a Venus Fly-trap mechanism. The asparate ligand was buried in the centre and made contacts with the two lobes by interactions of Ser164, Arg169 and Asp250 with the amine and carboxyl groups of the aspartate ligand. Arg147, Trp232 and Thr162 interacted with the carboxyl side chain of the aspartate ligand by hydrogen bonding and Lys161 contacted the aspartate ligand *via* a single water molecule. *Mtb* GlnH was also co-crystallised with lower-affinity amino acids. Interestingly the authors proposed that *Mtb* GlnH senses amino acid concentrations in the periplasm in order to adjust cellular nutrient responses to availability (Bhattacharyya et al., 2018).

The structure of *Mtb* UgpB was previously solved by X-ray crystallography to 1.5 Å resolution in the open ligand unbound conformation (Figure 5.2) (Jiang et al., 2014). As the ligand was not bound the molecular mechanisms of binding are still not known. As there are no structures available that have high homology to *Mtb* UgpB additional investigations are required to understand the molecular basis of ligand binding. From the previous study (Jiang et al., 2014) *Mtb* UgpB was identified to bind to glycerol-3-phosphocholine (GPC) but not glycerol-3-phosphate (G3P) by isothermal calorimetry however the authors were not able to obtain co-crystals of *Mtb* UgpB bound to either GPC or G3P. In comparison, a structure of G3P bound *E. coli* UgpB was previously solved by X-ray crystallography to 1.8 Å resolution (Wuttge et al., 2012) and it revealed that G3P interacted with the residues Trp169 and Trp172 *via* the glycerol backbone. Further stabilisation is provided by hydrogen bonding between Arg374, Glu66 that interact with the 1-hydroxyl and 2-hydroxyl of the glycerol group and Ser247, Tyr42 and Tyr323 that hydrogen bond with the phosphate group. The structures of GPC compared to G3P (Figure 5.3) indicate that both compounds have a glycerol backbone

and phosphate group however GPC also has a 2 additional CH_2 groups and a positively charged choline group which may be key determinants for binding to *Mtb* UgpB. We wanted to attempt to co-crystallise *Mtb* UgpB with GPC and the alternative glycerophosphodiesters shown (Figure 5.4).

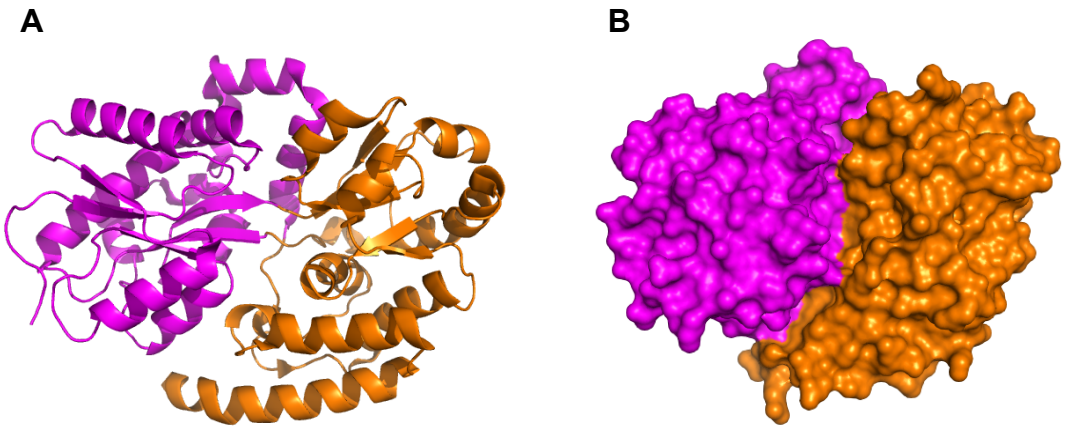


Figure 5.2: The X-ray crystal structure of ligand unbound apo *Mtb* UgpB (PDB 4MFI) (Jiang et al., 2014). A) Cartoon representation identifying the secondary structure elements domain I (magenta), domain II (orange). B) Surface representation where the two domains are highlighted, domain I (magenta) and domain II (orange).

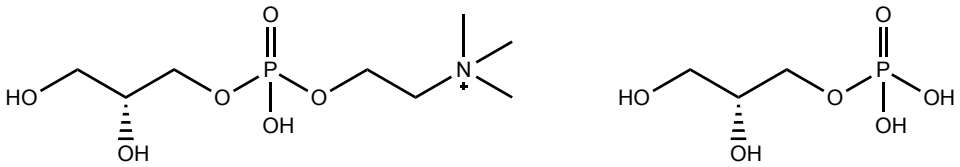


Figure 5.3: Structures of glycerol-3-phosphocholine (GPC) (left) and glycerol-3-phosphate (G3P) (right).

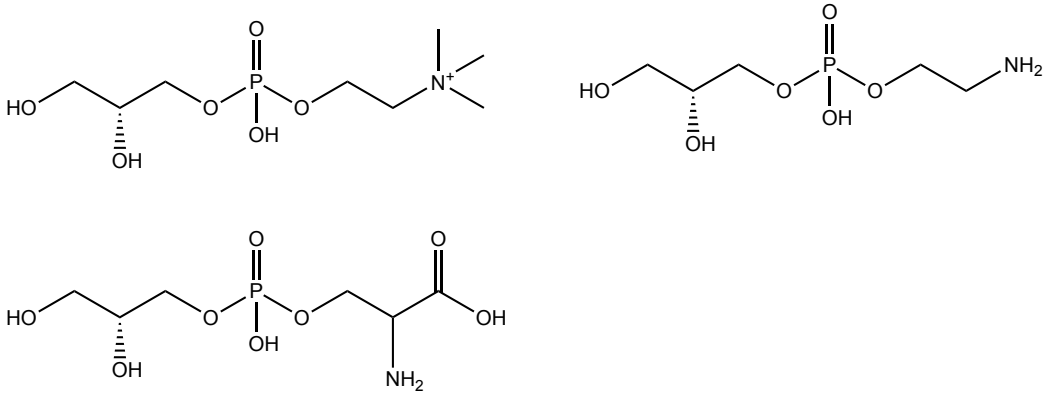


Figure 5.4: Structures of compounds to co-crystallise with *Mtb* UgpB. From left to right glycerol-3-phosphocholine (GPC), glycerophosphoethanolamine (GPE) and glycerophosphoserine (GPS).

5.2 Hypotheses and Aims

I hypothesised that the *Mtb* UgpB substrate-binding protein could be co-crystallised with GPC and alternative glycerophosphodiesters and the structures could be solved by X-ray crystallography. I also hypothesised that an in-solution method using STD NMR may reveal additional insights into the specific ligand binding recognition mode that could not be found in the solid state. As these potential ligands have a glycerol tail and phosphate group I hypothesised that these groups were the main ligand recognition sites, however that the head-group may also have a role in determining the affinity of the binding site interaction.

The specific aims of the work presented in this chapter were:

- To determine the X-ray co-crystal structure of GPC bound to *Mtb* UgpB.
- To determine the X-ray co-crystal structure of alternative glycerophosphodiesters bound to *Mtb* UgpB.
- To investigate *Mtb* UgpB ligand interactions in the solution state by STD-NMR and DEEP STD-NMR.

5.3 Results

5.3.1 Protein production

Mtb UgpB protein was produced as described (Chapter 3, 3.3.5 and 3.3.7). The protein was concentrated by ultrafiltration (10 kDa cut-off, Amicon Ultra) to 7.4 mg/mL. It is usually preferred to remove the hexa-his tag prior to setting down crystal trials (Majorek et al., 2014) however in the UgpB-pYUB1062 construct there is no cleavage site between the C-terminus and the hexa-His tag therefore crystal trials were set with the hexa-His tag intact.

5.3.2 Crystallisation studies of *Mtb* UgpB

For co-crystallisation experiments *Mtb* UgpB (7.4 mg/mL) and 10-100 mM of GPC were first incubated at 4 °C for 30 minutes. Crystal trials were set by using the sitting-drop vapour diffusion method in 96-well plates and a Mosquito liquid handling system (TTP LabTech). The crystallisation screens JCSG-*plus* and Morpheus were used for initial screening with drop sizes varying from 0.3-0.45 μL. Ratiometric volumes of 1:1, 1:2 and 2:1 volumes of protein with reservoir solution were mixed using the Mosquito liquid handling system. From these screening conditions, no crystals of *Mtb* UgpB formed. A previous study (Jiang et al., 2014) also attempted to co-crystallise GPC and *Mtb* UgpB however they were unsuccessful. Other attempts in the lab previously resulted in crystals of *Mtb* UgpB however they were always in the open conformation without ligand bound.

5.3.3 Chemical methylation of *Mtb* UgpB

In order to enhance co-crystallisation of GPC with *Mtb* UgpB we modified *Mtb* UgpB by reductive methylation using formaldehyde and dimethyamine borane complex (DMAB) (Walter et al., 2006). Reductive methylation of the surface-exposed lysine residues is used to improve the ability for proteins to crystallise because methylation of lysine residues alters the hydropathy, solubility and pI of the protein. These changes alter interactions between the protein-protein, protein-solvent and crystal lattice interfaces and in most cases the biological function is not affected (Walter et al., 2006). Following the methylation procedure crystal trials were repeated using the same concentrations and ratios (5.3.2) and the commercially available JCSG-*plus* screen. *Mtb* UgpB crystals typically grew within three days at 22 °C in the 1:1 ratiometric volume of protein with reservoir solution in the condition: 0.2 M $MgCl_2$, 0.1 M Tris-HCl pH 8.5, 20% w/v PEG 8,000. The crystals appearance was plate-like (Figure 5.5).

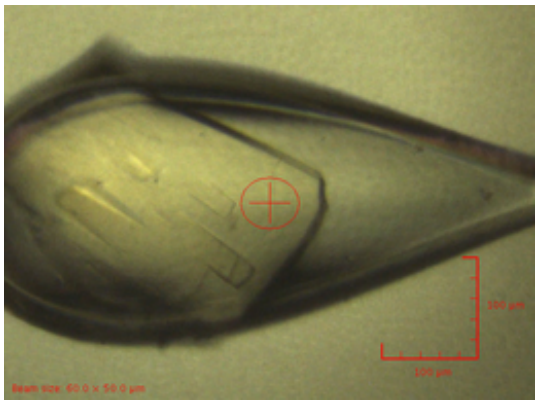


Figure 5.5: *Mtb* UgpB co-crystallised with glycerol-3-phosphocholine (GPC). Crystals formed with methylated *Mtb* UgpB and 10 mM GPC in the JCSG-*plus* screen and typically grew within three days at 22 °C in the condition 0.2 M $MgCl_2$, 0.1 M Tris-HCl pH 8.5, 20% w/v PEG 8,000.

5.3.4 Crystal structure of *Mtb* UgpB bound to glycerol-3-phosphocholine

The *Mtb* UgpB crystals that formed were cryoprotected with 20 % glycerol to reduce ice crystal formation (Pflugrath, 2015) and flash frozen in liquid nitrogen prior to diffraction data collection at the I04 beamline (Diamond Light Source). The diffraction data were indexed, integrated and scaled with XDS (Kabsch, 2010) through the XIA2 pipeline and the CCP4 suite of programmes (Collaborative Computational Project, 1994). Initially the ligand unbound *Mtb* UgpB structure (PDB ID 4MFI) was used as a search model to solve the phases by molecular replacement, however this failed. As SBP's are known to be very flexible the two-α/β domains were used as separate search models specifying (PDB ID 4MFI), residues 34-153 and 305-378 as domain I and residues 154-304 and 379-424 as domain II. A single molecule of 4MFI was searched for in the asymmetric unit however significant density was not fitted therefore the number of molecules searched for was increased. Four molecules were identified in the asymmetric unit with a Matthew's coefficient of 2.18 $Å^3$/Da and the space group was orthorhombic P 2_1 2_1 2_1. The structure was initially built using Autobuild (Terwilliger et al., 2008) followed by iterative cycles of alternating manual rebuilding in COOT (Emsley et al., 2004) and reciprocal space crystallographic refinement with PHENIX-REFINE (Afonine et al., 2012) assigning each domain as a separate TLS group. Residues 37-354 and 367-403 were clearly defined in all chains. However as residues 355-366 in chain C and D were not well defined they were deleted from the model. A single methylated lysine was confidently fitted in all chains at residue Lys161 out of a total of 13 lysine residues however due to the small size of the methyl group (15.03 g/mol) these were difficult to accurately fit as there was no clear density. In all four chains between the two-α/β domains, a central electron dense region was identified to be the bound ligand and glycerol-3-phosphocoline was manually fitted and the molecular restraints were calculated (eLBOW) (Adams et al., 2010). To finally refine

the model a single glycerol molecule was fitted in chains A and C and 2 glycerol molecules were fitted in chain B. A single Mg^{2+} ion was fitted in chains A, C and D and two Mg^{2+} ions were fitted in chain B. A total of 259 H_2O molecules were also fitted. All residues were found in the Ramachandran allowed regions (Davis et al., 2004) and no outliers were identified. The resolution of the final structure was 2.3 Å and the R_{work} was 20.6 % and R_{free} was 25.6 %. Crystallography parameters are listed (Table 5.1).

Table 5.1. Crystallographic parameters for *Mtb* UgpB in complex with GPC

PDB ID	**UgpB-GPC 6R1B**
Data collection	
Beam line	Diamond I04-1
Wavelength (Å)	0.92
Space group	P 2_1 2_1 2_1
Unit cell parameters	
a (Å)	169.9
b (Å)	213.3
c (Å)	46.1
α	90
β	90
γ	90
Molecules in ASU	4
Resolution (Å)	38-2.27
(Outer shell)[a]	(2.33-2.27)
Unique reflections	78,213 (5,626)
Multiplicity	7.8 (6.7)
$CC_{1/2}$	0.997 (0.476)
Completeness (%)[a]	99.2 (97.5)
R_{merge} (%)[a]	13.4 (5.1)
Mean I/σ(I)[a]	10.3 (1.7)
Refinement	
R work (%)	20.6
R free (%)	25.6
r.m.s.d	
Bond lengths (Å)	0.006
Bond angles (degrees)	0.66
No. of non-hydrogen atoms	
Protein atoms	11,953
Ligand/Ions	13
Solvent waters	259
Average B factors (Å^2)	
Overall	54.2
Protein	54.5
Ligand/Ions	42.7
Solvent	45.0
Ramachandran plot[b]	
Favoured region (%)	96.3
Allowed region (%)	3.72
Outer region (%)	0

[a]Numbers in parentheses refer to the highest-resolution shell.
[b]Ramachandran plot statistics were calculated by MolProbity.

5.3.5 Comparison of the four chains of *Mtb* UgpB

The four molecules of *Mtb* UgpB were aligned using PDBeFOLD (Hayward et al., 2002) and found an r.m.s.d of 0.35-0.44 Å for 394-395 residues indicating that the molecules in the asymmetric unit were equivalent (Figure 5.6A). PDBePISA (Krissinel et al., 2007) was used to analyse crystal packing interfaces and did not suggest that *Mtb* UgpB forms dimers or larger oligomers and is therefore consistent with substrate binding proteins of other ABC-transporters that also form monomers (Locher, 2009) (Scheepers et al., 2016). This also supports the size-exclusion chromatography studies (Chapter 3, 3.3.7) that demonstrate *Mtb* UgpB exists in solution as a monomer and the apparent molecular weight is 44 kDa therefore the physiologically relevant unit is most likely a monomer.

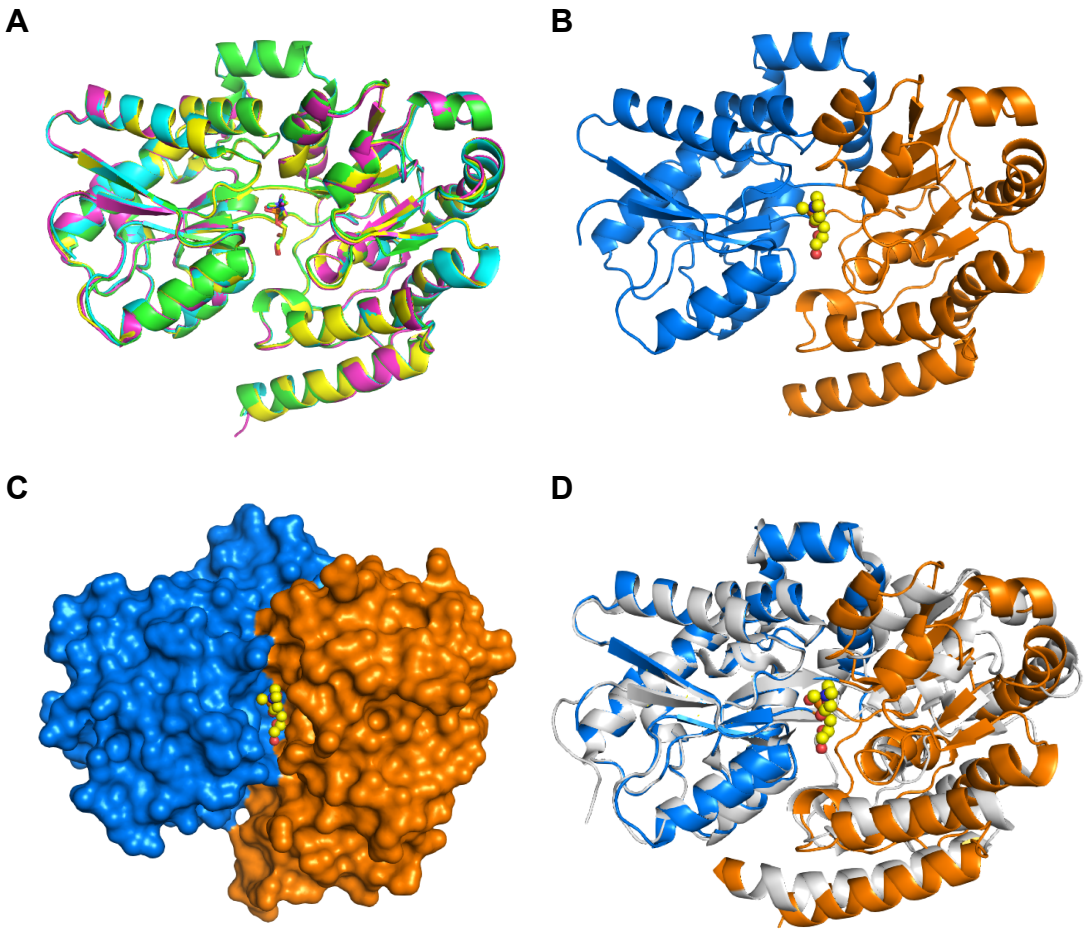


Figure 5.6: Overall structure of *Mtb* UgpB in complex with GPC. A) Alignment of the 4 chains of *Mtb* UgpB bound to glycerol-3-phosphocholine in the asymmetric unit. B) Cartoon representation identifying the secondary structure elements domain I (blue), domain II (orange). The GPC ligand is represented as spheres with yellow carbon atoms. C) Surface representation where the two domains are highlighted, domain I (blue) and domain II (orange). The GPC ligand is represented as spheres with yellow carbon atoms. D) Superposition of domain I of GPC *Mtb* UgpB co-complex (blue/orange) with domain I of apo *Mtb* UgpB (PDB 4MFI) (grey).

5.3.6 Overall structure of *Mtb* UgpB bound to glycerol-3-phosphocholine

The overall structure of *Mtb* UgpB (Figure 5.6B and 5.6C) consists of two-α/β domains connected by two flexible hinges between the residues Arg152-Pro155 and Ala290-307. Domain I is made of a five-stranded β-sheet (residues 1-154 and 307-365) surrounded by 11 α-helices whereas domain II is made of a four-stranded β-sheet that's enclosed by 9 α-helices. This classifies *Mtb* UgpB as a class II cluster D substrate binding protein (Berntsson et al., 2010) (Fukami-Kobayashi et al., 1999). The structure of ligand bound *Mtb* UgpB was compared to the unbound structure (PDB ID 4MFI) and revealed a 21.8 ^{o} rotation of domain I relative to domain II about the interdomain screw-axis from DynDom analysis (Table 5.2) (Hayward et al., 2002). Hinge-bending residues were identified as residues 152-153, 304-306 found in the two inter-domain connecting regions and also from residues 362-372 that are separate to the two main flexible hinges (Figure 5.7) (Table 5.2).

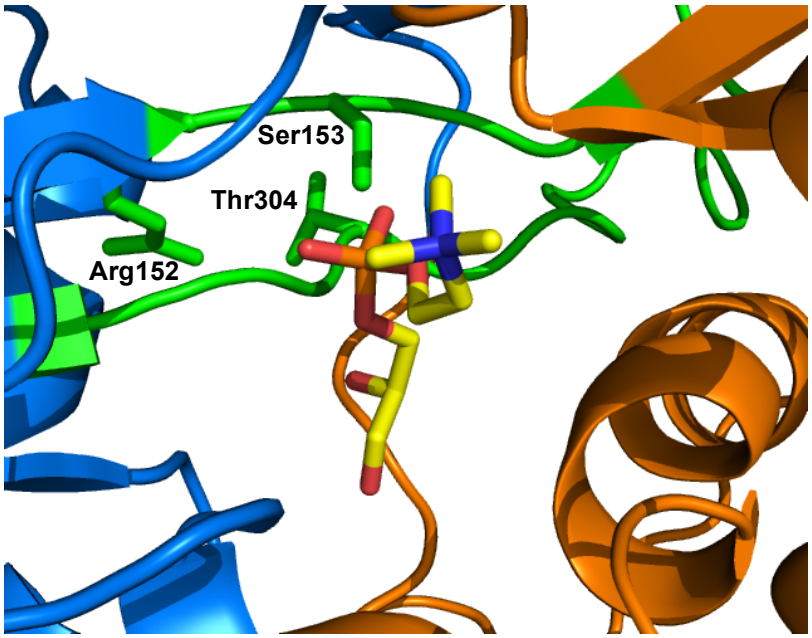


Figure 5.7: Flexible hinges and hinge-bending residues of *Mtb* UgpB. Flexible hinges between the two-α/β domains are residues Arg152-Pro155 and Ala290-Ala307 coloured green. The flexible hinge bending residues measured by DynDom analysis Arg152-Ser153 and Thr304-Gly306 are coloured as green sticks.

Table 5.2. DynDom analysis of *Mtb* UgpB (pdb 4MFI) and *Mtb* UgpB in complex with GPC

Backbone RMSD (Å)	Bending region	Rotation angle	Translation (Å)	Closure (%)
0.47 (Domain I)	152-153	21.8^{o}	0.8	98.7
0.54 (Domain II)	304-306			
	362-372			

Domain I comprises residues 38-152 and 306-365 and Domain II comprises residues 153-305 and 366-426

Mtb UgpB appeared to follow the Venus Fly-trap mechanism where the substrate binding protein closes upon ligand binding (Locher, 2009) (Scheepers et al., 2016). Ligand binding resulted in a bending movement that reduced the binding pocket volume almost two-fold from 1986 Å^3 to 791 Å^3 as measured by CAVER (Figure 5.8) (Kozlikova et al., 2014).

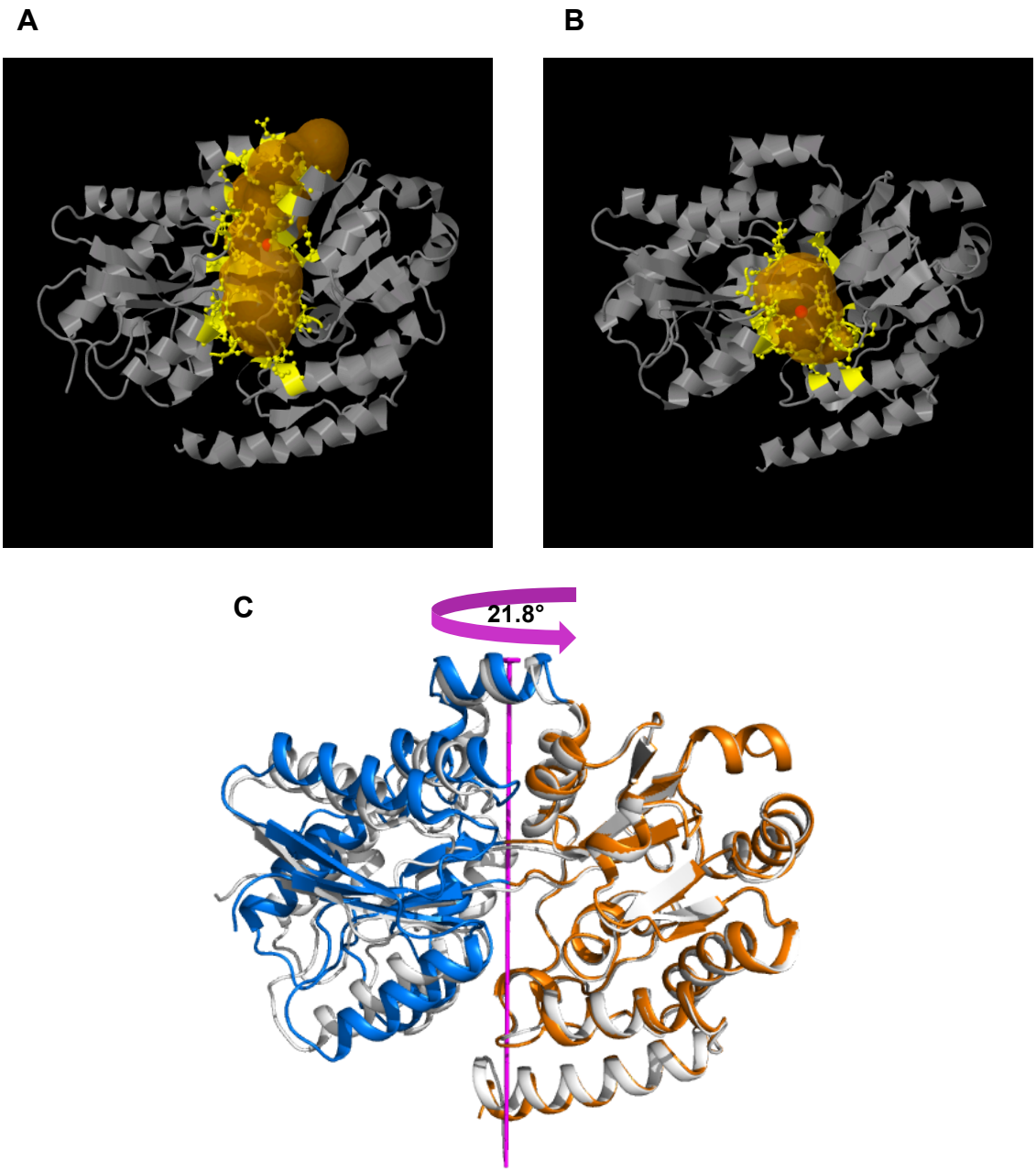


Figure 5.8: Binding pocket cavity volume of *Mtb* UgpB as determined by Caver. A) The measured binding pocket cavity volume of apo *Mtb* UgpB was 1986 Å^3. B) The measured binding pocket cavity volume of glycerol-3-phosphocholine bound *Mtb* UgpB was 791 Å^3. C) Rotation angle of 21.8 ^{o} between the domains of glycerol-3-phosphocholine bound *Mtb* UgpB (blue and orange) compared to apo *Mtb* UgpB (white), rotational axis (magenta).

The closed ligand bound conformation appeared to be stabilised by Arg385 of domain II forming interdomain hydrogen bonds with Asp102 of domain I and Gln381 from domain II. Alignment of GPC bound *Mtb* UgpB to the unbound structure (PDB ID 4MFI) (PDBeFOLD) found an r.m.s.d of 2.2 Å (over 385 residues) whereas alignment of the individual domains of the ligand bound structure and unbound structure (Figure 5.6D) found an r.m.s.d of 0.57 Å and 0.75 Å for domains I and II respectively (over 178 atoms (domain I) and over 216 atoms (domain II) (PDBeFOLD) (Hayward et al., 2002) indicating the importance of interdomain flexibility for ligand binding.

5.3.7 The glycerol-3-phosphocholine binding site in *Mtb* UgpB

The GPC ligand was manually fitted into the central electron dense region between the two-α/β domains identified in all four molecules of the asymmetric unit (Figure 5.9A). The GPC electron density map was contoured at 0.38 electrons/Å^3 (CCP4mg). GPC from chains A-D were aligned (PyMOL) and appeared to be orientated identically in each molecule (Figure 5.9B).

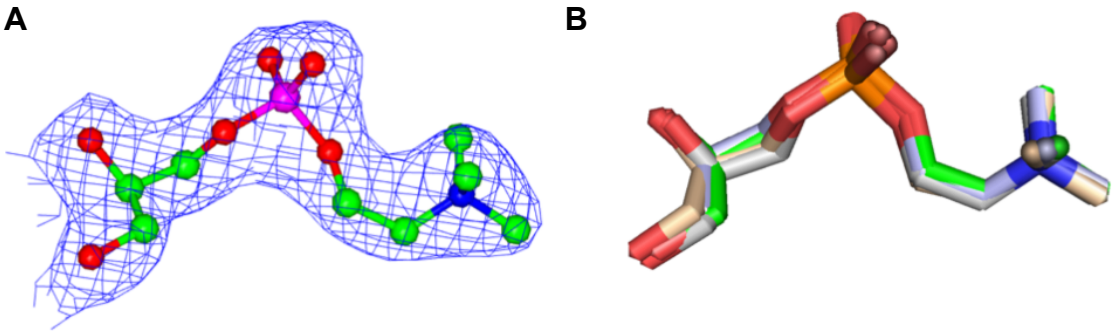


Figure 5.9: Glycerol-3-phosphocholine electron density. A) Electron density for the GPC substrate. Electron density map contoured at 0.38 electrons/Å^3. Carbon atoms are shown in green, oxygen atoms are shown in red, nitrogen atoms in blue and phosphate atom in purple. The figure was prepared using CCP4mg. The mtz file was loaded directly with the default settings and clipped to select for the GPC atoms. B) Alignment of the glycerophosphocholine ligand from each *Mtb* UgpB subunit. Superposition of GPC from each *Mtb* UgpB subunit in the asymmetric unit. The GPC ligand is shown with green carbon atoms (subunit A), light blue carbon atoms (subunit B), wheat carbon atoms (subunit C) and grey carbon atoms (subunit D). In all subunits oxygen atoms are coloured red and the phosphorous atom is coloured orange.

5.3.7.1 The electrostatic surface of GPC bound *Mtb* UgpB

The electrostatic surface of *Mtb* UgpB (CCP4mg) (Figure 5.10A) was analysed because it could be a determinant for ligand binding. The electrostatic surface indicated an acidic interface between the two-α/β domains of *Mtb* UgpB. The glycerol moiety of GPC is located at the base of this acidic cavity in close proximity to Arg385 and appears to stabilise domain closure whereas the choline moiety is pointing outwards and is exposed to solvent (Figure 5.10B).

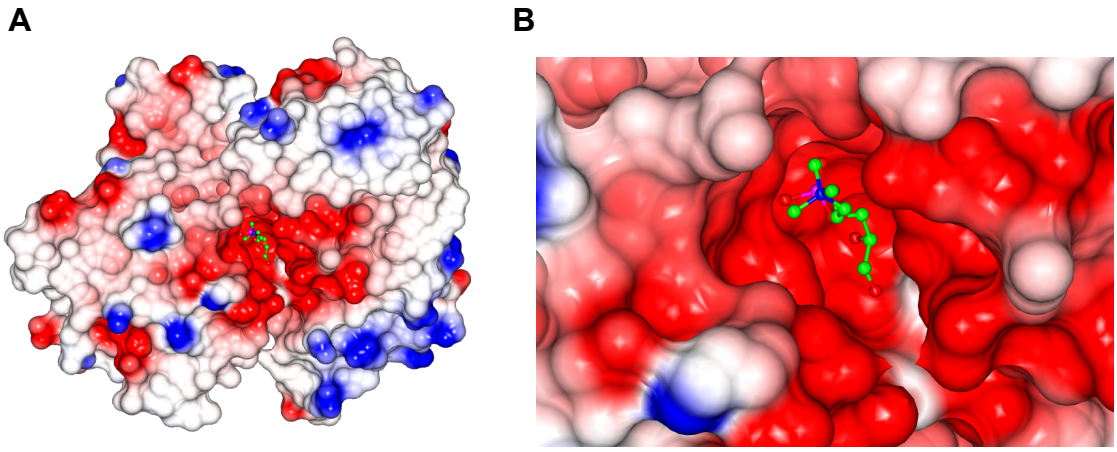


Figure 5.10: The binding pocket electrostatic surface of *Mtb* UgpB (CCPMG). A) Electrostatic surface representation of *Mtb* UgpB GPC. B) Close-up illustration of the *Mtb* UgpB binding pocket with the GPC ligand in stick representation. Electrostatic surface was generated in CCP4mg (Collaborative Computational Project, 1994) red represents negative potential, blue represents positive potential.

5.3.7.2 Ligand binding site interactions of *Mtb* UgpB

The interactions of the GPC ligand with *Mtb* UgpB were now investigated in detail. The side chains of Leu205 and Trp208 from domain II form hydrophobic contacts with the glycerol moiety (Figure 5.11A and 5.11B). Parallel to the plane formed by these atoms: C1, C2 and the 2-hydroxy group of the glycerol moiety is Trp208 where its indole ring system enables π-stacking interactions. Perpendicular to the same plane there is additional stabilisation provided by the aliphatic residue, Leu205 although it is approximately 4 Å distance away from the glycerol backbone. Within the binding pocket multiple hydrogen bonding interactions provided by 7 additional residues fix the GPC ligand in place. From domain I, direct hydrogen bonding occurs between Asp102 and the 1- and 2-hydroxyl groups of the glycerol moiety. Two residues of the flexible hinge region bind to the GPC ligand and may be important to stabilise the ligand bound closed conformation: Arg385 (domain II) hydrogen bonds to the 1-hydroxy of the glycerol moiety and Gly306 (domain I) also hydrogen bonds with the 2-hydroxy of the glycerol moiety via its backbone amide nitrogen atom. Further ligand binding interactions are stabilised *via* the phosphate moiety of GPC and hydrogen bond interactions between the side chains of Tyr78, Tyr345 and Ser153 (domain I) and the Ser272 (domain II) side chain and Gly306 amide backbone (domain I). Interestingly no interactions appeared to occur between *Mtb* UgpB and the positively charged choline moiety despite its well-defined density and the choline group being a key requirement for binding based on ligand binding studies (Chapter 4, 4.3.12). Therefore it could be combination of ligand charge and size that are required for recognition. Based on the

lack of interactions with the choline group, other ligands could potentially be recognised by *Mtb* UgpB that have a glycerol-3-phosphate core but different head-group.

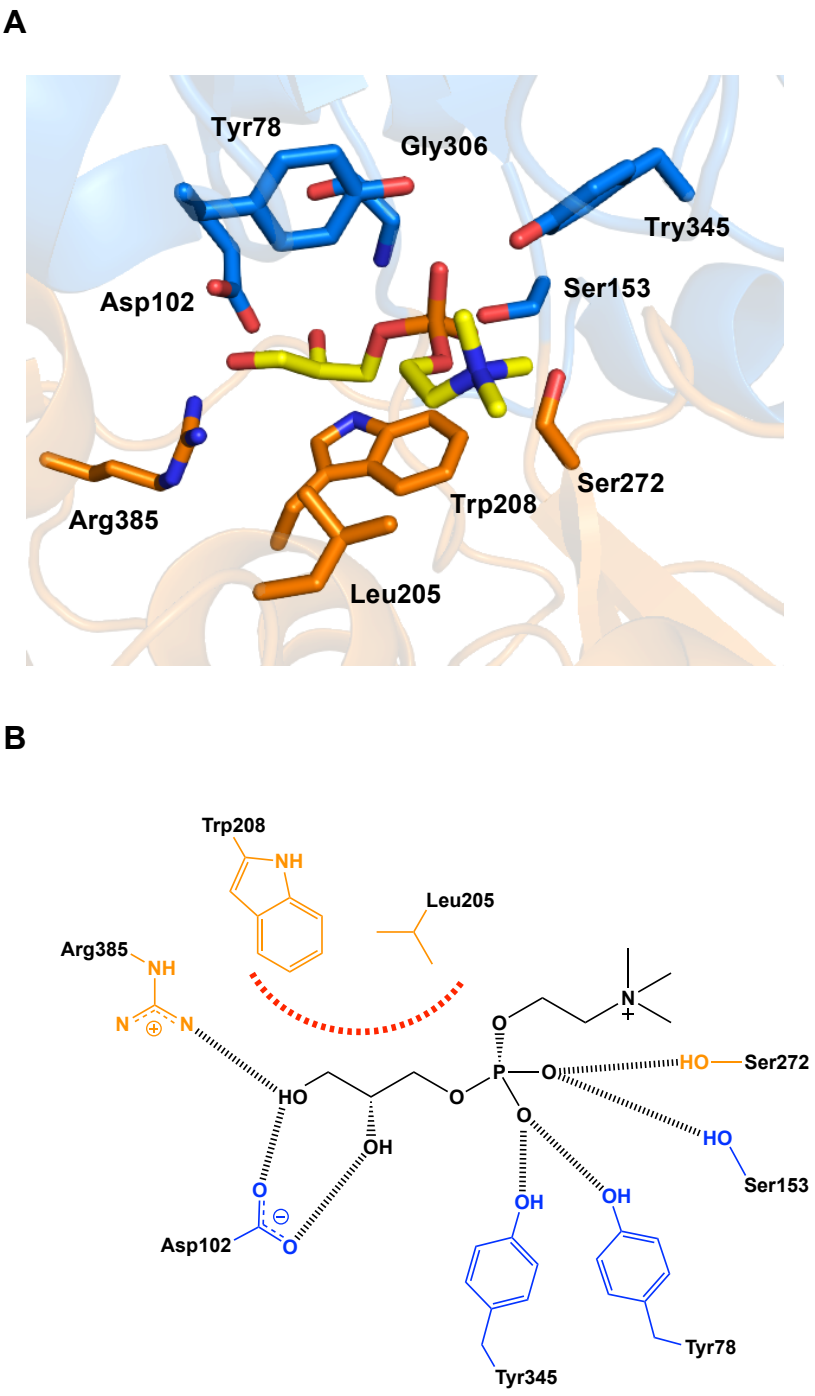


Figure 5.11: The GPC binding site in *Mtb* UgpB. A) Illustration showing GPC with yellow carbon atoms and selected *Mtb* UgpB amino acid residues in stick representation (coloured blue for residues within domain I, and orange for residues within domain II. B) Schematic diagram of the interactions of *Mtb* UgpB with GPC. Dashed lines (black) represent hydrogen bonding, thick dotted line (red) represents hydrophobic interactions.

5.3.8 Identification of an additional glycerol in the *Mtb* UgpB binding site

Interestingly in chain B of GPC bound *Mtb* UgpB an additional glycerol was identified (Figure 5.12A). This could be a result of cryo-protecting the crystals with glycerol and also because glycerol is included in the purification buffer. The glycerol molecule was found in the *Mtb* UgpB binding pocket and within 4 Å close proximity to the choline moiety of the GPC ligand (Figure 5.12B). This is similar to the G3P bound *E. coli* UgpB structure (Wuttge et al., 2012) where a glycerol molecule is also present in the binding pocket although at a different position and therefore indicates that the binding pockets of both UgpB proteins are larger than the GPC ligand. This may be functionally important for accommodating larger glycerophosphodiester ligands.

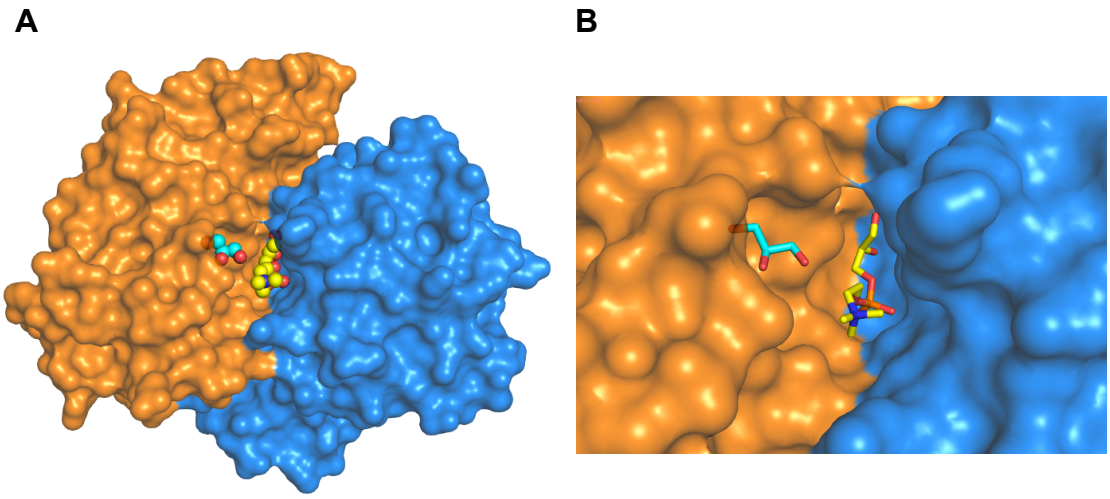


Figure 5.12: Location of additional glycerol moiety in the *Mtb* UgpB binding pocket. A) Surface representation of *Mtb* UgpB. The GPC ligand is represented by yellow spheres and a solvent glycerol moiety as cyan spheres. B) Close-up of the *Mtb* UgpB binding pocket. The GPC ligand and glycerol moiety are shown in stick representation.

5.3.9 Comparison of *Mtb* UgpB and *E. coli* UgpB

The only other available structure of a UgpB substrate-binding protein is from *E. coli* and it is bound to glycerol-3-phosphate (G3P). The binding profiles of *E.coli* UgpB and *Mtb* UgpB are different because *Mtb* UgpB did not show binding to G3P by microscale themophoresis (Chapter 4, 4.3.12) or by isothermal calorimetry (Jiang et al., 2014). However when the overall structures of G3P bound *E. coli* UgpB (Wuttge et al., 2012) and GPC bound *Mtb* UgpB were compared they were found to be highly similar with a r.m.s.d. of 2.1 Å (PDBeFold, target residues: 394, sequence identity 25 % (Figure 5.13A) (PDB ID 4AQ4) (Krissinel et al., 2007). It appears that despite the inability of *Mtb* UgpB to recognise the smaller G3P ligand, the binding mode of the G3P core of GPC resembles the G3P *E. coli* UgpB bound structure (Wuttge et al., 2012) and the ligand-binding pocket of *Mtb* UgpB is very similar to *E. coli* UgpB (Figure 5.13B and

5.13D). However despite these overall similarities, important differences exist between the two UgpB structures. In *Mtb* UgpB Leu205 is replaced by the larger indole side chain of Trp269 in *E. coli* UgpB and Asp102 in *Mtb* UgpB is replaced by the longer and more acidic side chain of Glu66 in *E. coli* UgpB. These different residues may influence the different ligand selectivity. The two substrate-binding proteins may have also evolutionary diverged because the conserved Arg385 in *Mtb* UgpB and Arg374 in *E. coli* UgpB are located in different regions (Figure 5.13B).

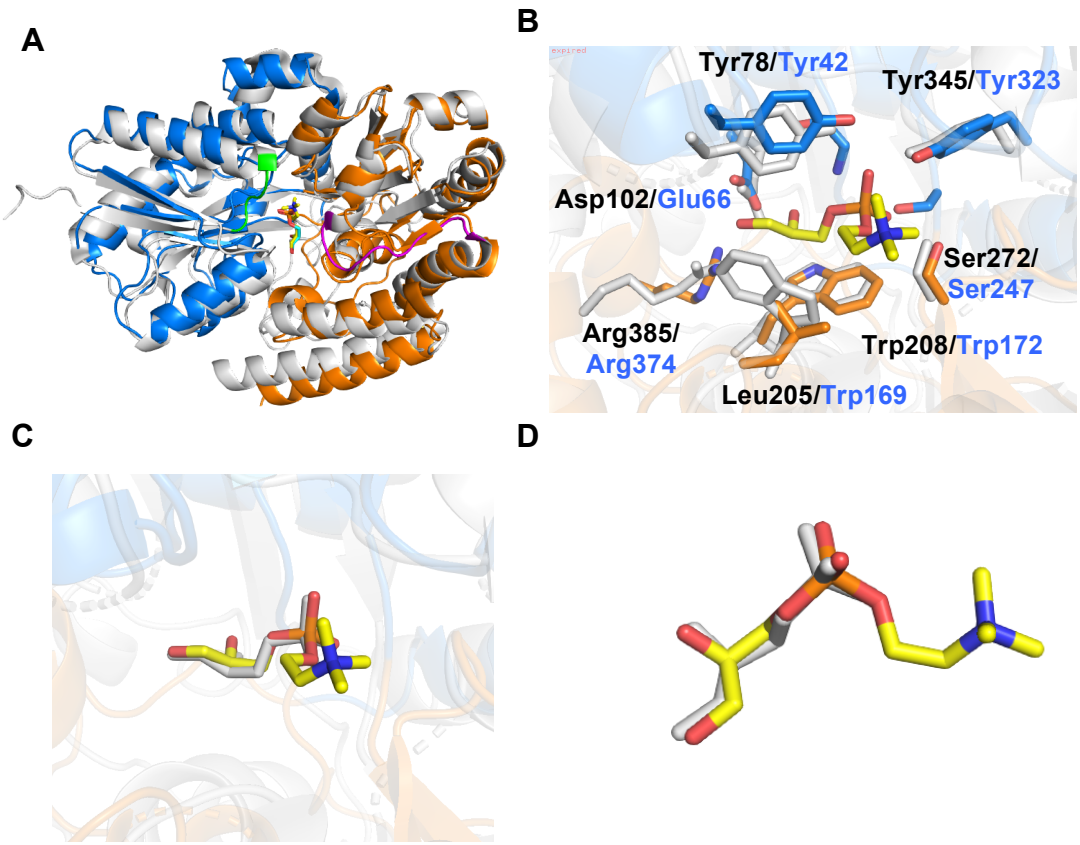


Figure 5.13: Comparison of glycerol-3-phosphocholine bound *Mtb* UgpB to glycerol-3-phosphate bound *E. coli* UgpB. A) Comparison of *Mtb* UgpB with *E. coli* UgpB. Superposition of the *Mtb* UgpB GPC complex structure (blue/orange) with *E. coli* UgpB in complex with G3P (PDB 4AQ4) (white). Loop regions that differ are highlighted in green and magenta. B) Close-up of the overlay of the binding-sites of GPC (*Mtb*) and G3P (*E. coli*). Selected residues are shown as sticks (*Mtb* blue/orange, *E. coli* white) and the font labeled in black (*Mtb*) and blue (*E. coli*). C) Close-up illustration showing the binding orientation of the GPC ligand and G3P ligand in stick representation (yellow carbon atoms, GPC, white carbon atoms G3P). D) Close up overlay of GPC and G3P.

Overall the substrate-binding pocket is much narrower in *E. coli* UgpB (Figure 5.14) compared to *Mtb* UgpB. This is partly due to two different loop regions in *E. coli* UgpB from residues Gly221-Asp230 (domain II) and residues Gly12-Pro30 (domain I). Gly221-Asp230 link α-helices 10 and 11 and cause a 5 Å translational shift that

narrows the substrate-binding pocket. A second loop His8-Gly12 causes the first α-helix (residues 12-30) to translate by 6 Å towards α-helix 11. Furthermore, compared to *E. coli* UgpB the entrance to the binding pocket is larger in *Mtb* UgpB.

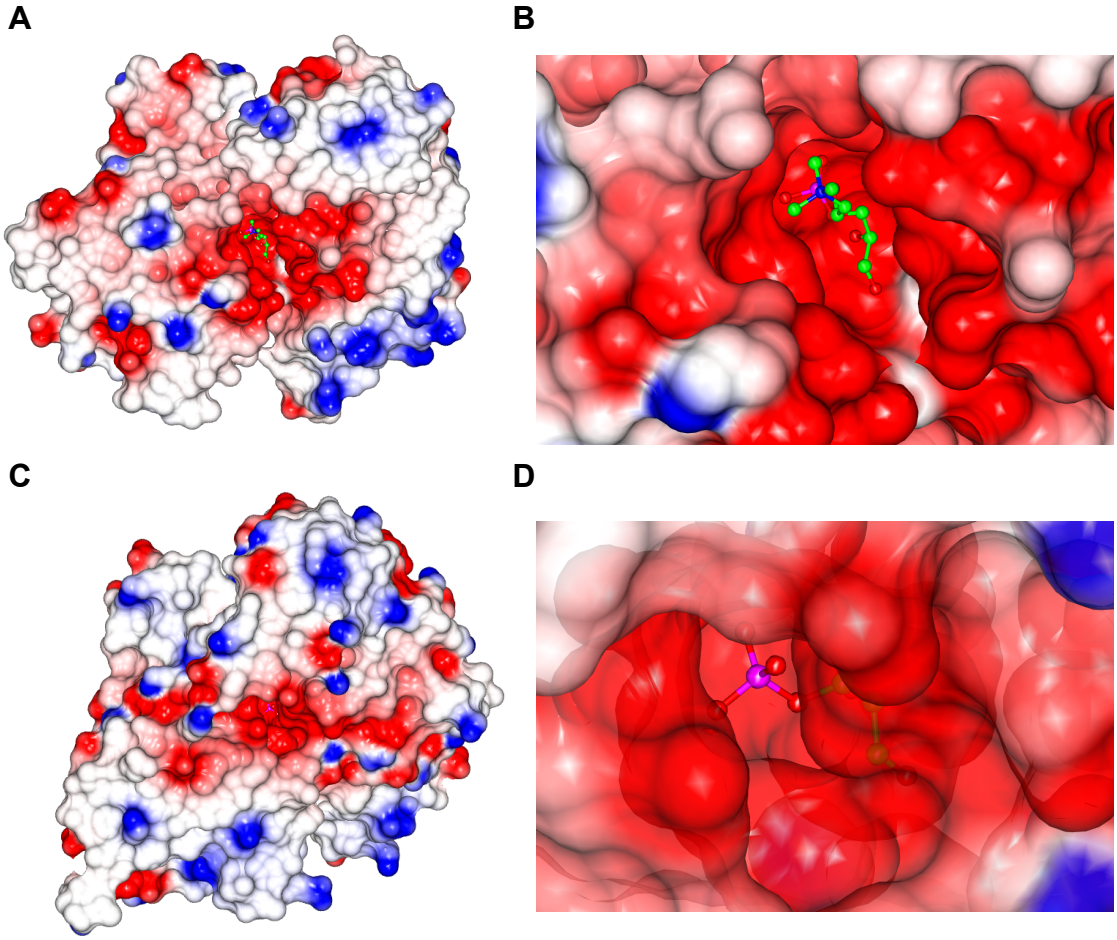


Figure 5.14: Comparison of the binding pocket electrostatic surface of *Mtb* UgpB and *E. coli* UgpB. A) Electrostatic surface representation of *Mtb* UgpB GPC. B) Close-up illustration of the *Mtb* UgpB binding pocket with the GPC ligand in stick representation. C) Electrostatic surface representation of *E. coli* UgpB G3P. D) Close-up illustration of the *E. coli* UgpB binding pocket with the G3P ligand in stick representation. Electrostatic surface was generated in CCP4mg (Collaborative Computational Project, 1994) red represents negative potential, blue represents positive potential.

5.3.10 Crystallisation studies of *Mtb* UgpB and glycerophosphoserine (GPS)

Other glycerophosphodiesters have been identified as substrates of *Mtb* UgpB (Chapter 4) therefore we wanted to investigate how the binding recognition mode of these substrates compares to GPC, as they have different binding affinities. GPS was investigated because it has a similar binding affinity of 14.9 μM compared to the binding affinity for GPC of 3.6 μM. For co-crystallisation experiments *Mtb* UgpB was reductively methylated as described previously (5.3.2). 20 mM of glycerophosphoserine (GPS) was incubated with methylated *Mtb* UgpB (10 mg/mL) at

4 °C for 30 minutes before setting plates. Crystal trials were set using the sitting-drop vapour diffusion method in 96-well plates using a Mosquito liquid handling system (TTP LabTech). Ratiometric volumes of 1:1, 1:2 and 2:1 volumes of protein with reservoir solution were mixed producing drop sizes of 0.3-0.45 μL. The crystallisation conditions that were initially screened included JCSG-*plus*, Morpheus, Midas and Structure Screen I and II (Molecular Dimensions). A single crystal was identified in the 1:1 ratio drop of JCSG-*plus* condition D4 (0.1 M sodium acetate, 0.2 M lithium sulphate, pH 4.5, 30 % w/v PEG8000) and diffracted at >4 Å resolution indicating that the crystal was protein however due to the low resolution further optimisation was required.

The Hampton additive screen (Hampton Research) was used at the manufacturers guidance to give a final 10 % additive concentration to condition D4 from JCSG-*plus* (0.1 M sodium acetate, 0.2 M lithium sulphate, pH 4.5, 30 % w/v PEG8000) that had previously produced a low diffraction quality protein crystal. In the additive screen several crystals were identified after 2 weeks in the additives: ethylene glycol (3.0 %), acetone (4.0 %), ethyl acetate (0.5 %) and trimethylamine hydrochloride (0.01 M). The crystals were flash frozen in liquid nitrogen prior to diffraction data collection at the I-24 beamline (Figure 5.15) (Diamond Light Source). No cryo-protection was added due to the high concentration of PEG8000 (30 % w/v) in the crystallisation condition. The crystal that produced high-resolution diffraction data (1.76 Å) grew in the additive ethylene glycol (3.0 %). Ethylene glycol is a polyol and has general structural stabilising properties as it increases the thermal denaturation temperature of proteins and may have induced order in the flexible *Mtb* UgpB protein to promote ordered crystal packing (Cudney et al., 1994).

Figure 5.15: *Mtb* UgpB co-crystallised with glycerophosphoserine (GPS). Crystals formed with methylated *Mtb* UgpB and 20 mM GPS in the JCSG-*plus* screen and Hampton additive. The crystal that diffracted was identified in the following condition: 0.1 M sodium acetate, 0.2 M lithium sulphate, pH 4.5, 30 % w/v PEG8000, 3.0 % ethylene glycol.

Table 5.3. Crystallographic parameters for *Mtb* UgpB in complex with GPS

	UgpB-GPS
PDB ID	**-**
Data collection	
Beam line	Diamond I-24
Wavelength (Å)	0.9687
Space group	P 1 2_1 1
Unit cell parameters	
a (Å)	54.6
b (Å)	46.7
c (Å)	80.3
α	90
β	109.2
γ	90
Molecules in ASU	1
Resolution (Å)	39.74-1.76
(Outer shell)[a]	(1.79-1.76)
Unique reflections	38,089 (3,684)
Multiplicity	5.4 (5.3)
$CC_{1/2}$	1.0 (0.4)
Completeness (%)[a]	99.8 (96.1)
R_{merge} (I)[a]	0.3 (1.9)
Mean I/σ(I)[a]	5.6 (1.6)
Refinement	
R work (%)	18.4
R free (%)	22.1
r.m.s.d	
Bond lengths (Å)	0.006
Bond angles (degrees)	0.82
No. of non-hydrogen atoms	
Protein atoms	3,016
Ligand/Ions	1
Solvent waters	284
Average B factors (Å^2)	
Overall	15.9
Protein	15.4
Ligand/Ions	18.9
Solvent	20.5
Ramachandran plot[b]	
Favoured region (%)	98.46
Allowed region (%)	1.28
Outer region (%)	0.26

[a]Numbers in parentheses refer to the highest-resolution shell.
[b]Ramachandran plot statistics were calculated by MolProbity.

5.3.11 Structure of *Mtb* UgpB co-crystallised with glycerophosphoserine

To solve the phases molecular replacement (PHASER) (McCoy et al., 2007) was used specifying chain A of GPC bound *Mtb* UgpB as the single search model. Only one molecule was identified in the asymmetric unit with a Matthew's coefficient of 2.2 Å^3/Da and the space group was monoclinic P12$_1$1. A full list of the crystallography parameters is presented (Table 5.3). The calculated phases were well defined and the model was

visualised (COOT) (Emsley et al., 2004) and appeared to be identical to GPC bound *Mtb* UgpB. However *Mtb* UgpB amino acids 90-93 did not fit well to the density and had to be manually adjusted. No methylated lysines were identified in the structure despite reductively methylating the protein prior to crystal trials. At the centre of the structure between the two-α/β domains an electron dense region was identified and we attempted to manually fit GPS. The glycerol and phosphate electron densities of GPS were clearly defined and GPS was manually fitted and the molecular restraints were calculated (eLBOW) (Adams et al., 2010). The putative GPS electron density map was contoured at 0.38 electrons/Å^3 (CCP4mg) (Figure 5.17). However the serine density of GPS was not well defined therefore we could not precisely confirm that GPS was bound and further studies are needed. In an attempt to improve the serine density we used different numbers of images and re-processed the data however the serine density did not improve. In addition, there was also an unknown density close to the putative GPS density. To finally build the model, H_2O molecules were fitted into the appropriate densities. The final overall structure (Figure 5.16) was 1.76 Å resolution and the R_{work} was 18.4 % and R_{free} was 22.1 %. A single molecule of *Mtb* UgpB, putatively co-crystallised with glycerophosphoserine (GPS) was identified in the asymmetric unit.

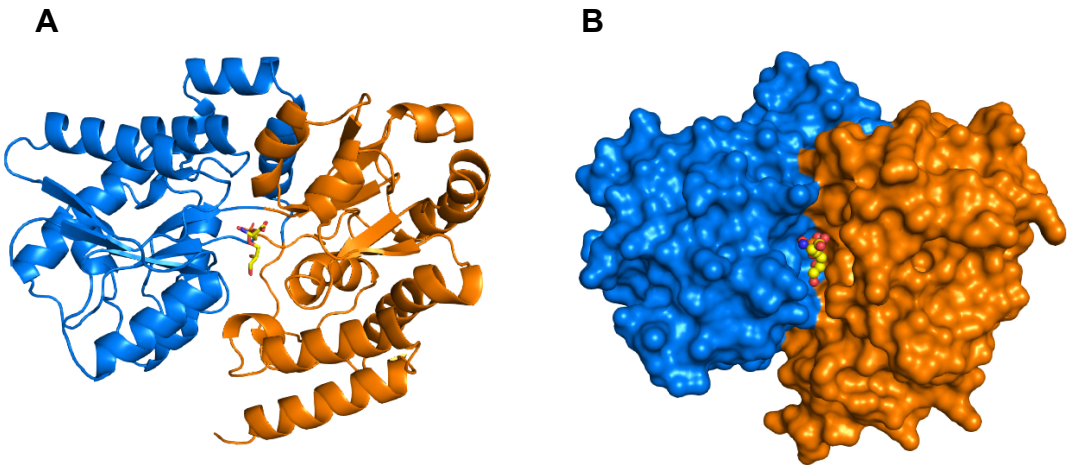


Figure 5.16: Overall structure of *Mtb* UgpB co-crystallised with GPS. A) Cartoon representation identifying the secondary structure elements domain I (blue), domain II (orange). The GPS ligand is represented as sticks with yellow carbon atoms. B) Surface representation where the two domains are highlighted, domain I (blue) and domain II (orange). The GPS ligand is represented as spheres with yellow carbon atoms.

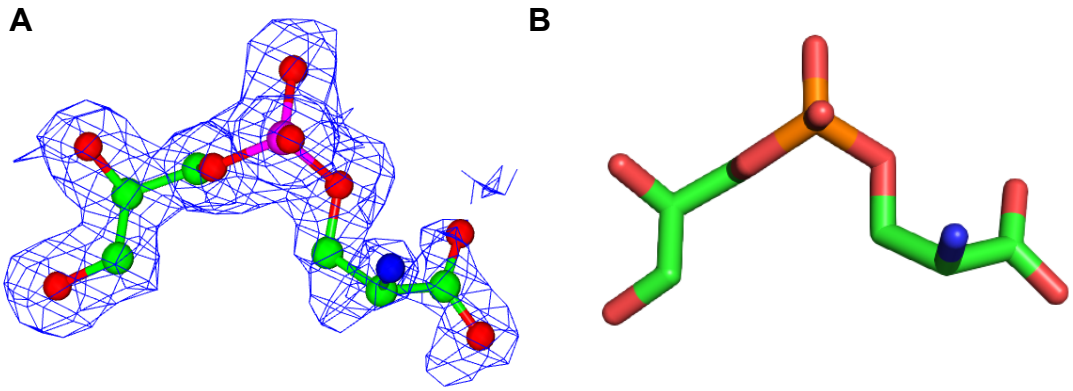


Figure 5.17: Putative glycerophosphoserine bound electron density. A) Electron density for the GPS ligand. Electron density map contoured at 0.38 electrons/Å^3. Carbon atoms are shown in green, oxygen atoms are shown in red, nitrogen atoms in blue and phosphate atom in purple. The figure was prepared using CCP4mg. The mtz file was loaded directly with the default settings and clipped to select for the GPC atoms. B) GPS ligand in stick representation showing green carbon atoms, red oxygen atoms, blue nitrogen atoms and orange phosphate atoms.

5.3.12 Comparison of *Mtb* UgpB bound GPC with GPS

Even though we were not completely sure if GPS is bound we compared the GPC bound *Mtb* UgpB structure with the putative GPS bound *Mtb* UgpB structure and found that they were almost identical (RMSD 0.335 over 392 residues). Comparisons of the residues involved in binding (Figure 5.18) revealed that the same residues were involved in binding and that they were orientated in the same way as the GPC bound structure.

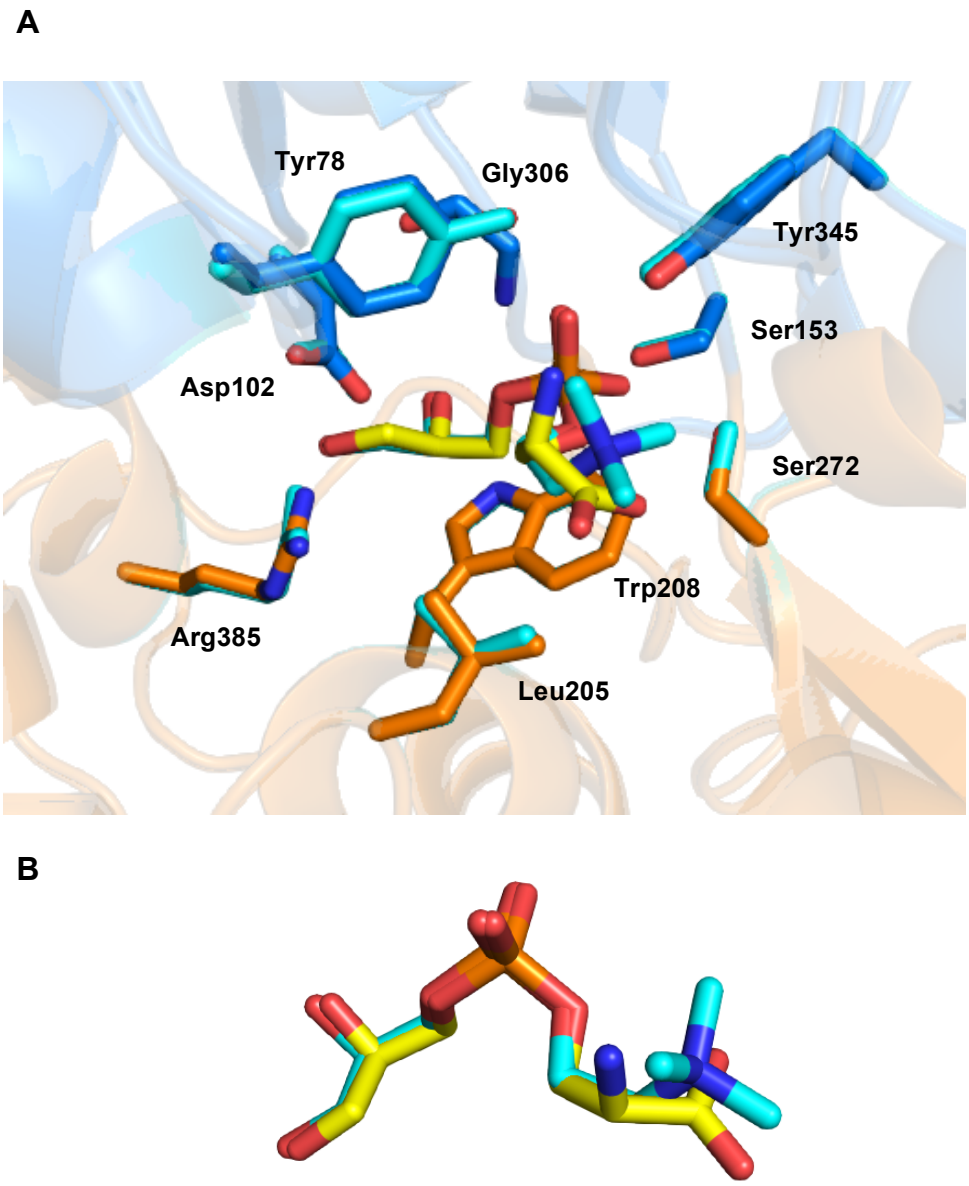


Figure 5.18: Comparison of the putative GPS bound *Mtb* UgpB binding site to GPC bound *Mtb* UgpB. A) Comparison of the putative GPS bound *Mtb* UgpB binding site to GPC bound *Mtb* UgpB. Selected residues are shown as sticks (*Mtb* GPS orange/blue, *Mtb* GPC cyan). The GPS ligand carbon atoms are coloured yellow and the GPC ligand carbon atoms are coloured cyan. B) Close-up illustration showing the putative binding orientation of the GPS ligand and GPC ligand in stick representation (yellow carbon atoms, GPS, cyan carbon atoms GPC).

The serine density of GPS was not well defined. However, no *Mtb* UgpB interactions could be identified with the serine group similar to how no interactions were identified with the choline head-group of the GPC bound structure (Figure 5.18A). The same residues were found to interact with the glycerol and phosphate group of GPS. The glycerol and phosphate group of the putative GPS ligand aligned almost identically with GPC (Figure 5.18B). Similar to GPC bound *Mtb* UgpB the glycerol C1 -hydroxyl group of GPS forms hydrogen bond interactions with Asp102 and Arg385. The glycerol C2 -

hydroxyl group of GPS forms hydrogen bonds with Asp102 and the nitrogen amide backbone of Gly306. Trp208 and Leu205 also stabilise the glycerol backbone via π-stacking interactions and Van der Waal's forces. Tyr78, Tyr345, Ser153 and Ser272 form hydrogen bonds with the phosphate hydroxyl groups of GPS. Asp102 (domain I) and Arg385 (domain II) were also identified to form a salt bridge that may be important for stabilising the closed conformation of the ligand bound structure and may be involved in enabling larger glycerophosphodiester ligands to bind.

5.3.13 STD-NMR for *Mtb* UgpB with glycerol-3-phosphocholine

Our collaborators Dr Ridvan Nepravishta and Dr Jesus Angulo at the University of East Anglia carried out all of the following STD-NMR experiments and analyses.

As no interactions were identified by X-ray diffraction between the choline moiety of GPC and *Mtb* UgpB, and G3P did not show binding despite the lack of a choline moiety, the binding interaction was investigated using NMR. As this method is in the solution state, it's anticipated that the protein is more conformational flexible and dynamic which may be important prior to contact with the ligand. Whereas protein crystals are in the solid state and tend to crystallise in a single conformation and are more dynamically restricted. To obtain quantitative epitope maps of the ligand-protein interaction in solution, STD-NMR was used (Mayer et al., 1999). Firstly, *Mtb* UgpB was buffer exchanged into deuterated PBS at pH 7.5 by ultrafiltration (10 kDa cut-off, Amicon Ultra) to a final concentration of 3.1 mg/mL (68 μM) and the ligands were also dissolved in deuterated PBS at 5 mM prior to the STD-NMR analyses.

In the STD-NMR experiment binding was successfully detected to GPC and the STD NMR signals and GPC binding epitope maps were obtained (Figure 5.19A). The epitope map clearly indicated that the glycerol moiety of GPC was the main recognition site as it showed the highest normalised STD values. Of the protons identified the highest STD intensities were found in the glycerol moiety at position 2 and 3 (H2G, 100 % and H3G, 100 %) (Figure 5.19A) whereas the proton at position 1 (H1G, 82 %) showed a slightly lower signal. The ligand-protein contacts appeared to be closer to the glycerol group compared to the choline group because the STD signals decreased from the glycerol group to the choline group. Protons at position 1 and 2 adjacent to the phosphate and choline group (H1C, 70 % and H2C, 50 %) showed intermediate STD signals and the lowest STD signals were observed for the methyl groups of the choline moiety (H3C, 26 %).

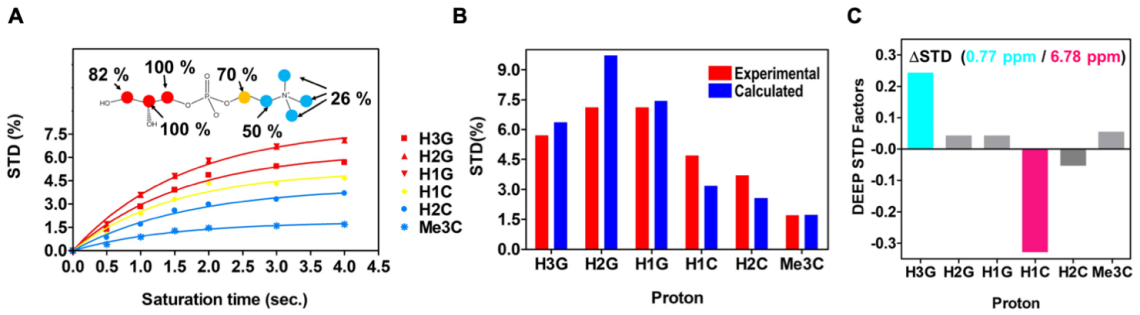


Figure 5.19: STD-NMR for *Mtb* with GPC. A) Structure of glycerol-3-phosphocholine. B) Experimental STD build up curves for the GPC/*Mtb* UgpB complex and the obtained epitope map of GPC/*Mtb* UgpB. C) STD in red bars obtained with a 4s saturation time while in blue bars the CORCEMA-ST calculated STD from the 3D crystallographic structure of the *Mtb* UgpB/GPC complex obtained for the same saturation time. RNOE factor 0.25. D) Differential epitope (DEEP)-STD factors showing the type of amino acid that the protons of the GPC ligand are orientated towards. Protons orientated towards aliphatic residues are highlighted in blue and protons orientated towards aromatic residues are highlighted in magenta. The figure was prepared by Dr Ridvan Nepravishta.

As the X-ray structure of ligand unbound *Mtb* UgpB was available a quantitative comparison of the NMR solution data and X-ray data was carried out by complete relaxation and conformational exchange matrix (CORCEMA-ST) calculations (Jayalakshmi et al., 2002) and the newly developed method DEEP-STD NMR (Monaco et al., 2017). Comparing the CORCEMA-ST calculated solution data STD NMR intensities (Figure 5.19B) to the X-ray structure revealed a NOE R-factor (Xu et al., 1995) of 0.25 and demonstrates very good data agreement as a NOE R-factor of 0.31-0.39 is typical for accurate structures (Edmondson, 1992). However to probe for additional information about how the ligand is orientated in the binding site as well as indirect information about the types of amino acids (aromatic, polar or apolar residues) contacting the ligand, differential epitope mapping STD NMR (DEEP-STD NMR) was used (Figure 5.19C) (Nepravishta et al., 2019). The calculated DEEP-STD NMR factors indicated that protons in position 3 of the glycerol moiety of GPC (H3G) are orientated towards aliphatic amino acids whilst the protons in position 1 adjacent to the phosphate and choline moiety (H1C) are oriented toward aromatic residues and these were mapped to the crystal structure of *Mtb* UgpB as the aliphatic Leu205 and aromatic Tyr78 and Tyr345 amino acid residues. This data correlates well with the ligand recognition mode obtained for the GPC *Mtb* UgpB complex (Figure 5.11) in the solid-state structure showing the side chains of Leu205 from domain II forming hydrophobic contacts with the glycerol moiety and Tyr78 and Tyr345 from domain I forming hydrogen bond interactions with the phosphate moiety next to proton (H1C).

5.3.14 STD-NMR for *Mtb* UgpB with glycerolphosphoinositol-4-phosphate

As binding to the alternative glycerophosphodiester, glycerophosphoinositol-4-phosphate (GPI(4)P) was identified by microscale thermophoresis with a K_d of 298.8 μM (Chapter 4, 4.5.1), STD-NMR spectroscopy was used to investigate the binding interaction with *Mtb* UgpB. Crystal trials of GPI(4)P and *Mtb* UgpB were not set because of its limited commercial availability and high purchase cost. The results indicated similarly that the glycerol moiety of GPI(4)P was the main recognition site as the strongest STD signals were at position 1, 2 and 3 (H1G, 93 % and 88.7 %), (H2G, 100 %) and (H3G, 82 % and 77 %). However also high STD signals were observed for protons located in the inositol ring at position 1 and 2 (H1I, 71%) and (H2I, 91 %) (Figure 5.20A). Intermediate STD signals were identified for protons located in the inositol ring at positions 3 and 4 (H3I, 65 %) and (H4I, 67 %) whereas protons at position 5 and 6 showed the lowest STD signals (H5I, 46 %) and (H6I, 42 %). This contrasts with GPC where only low STD signals were observed for the choline head-group (Figure 5.20A).

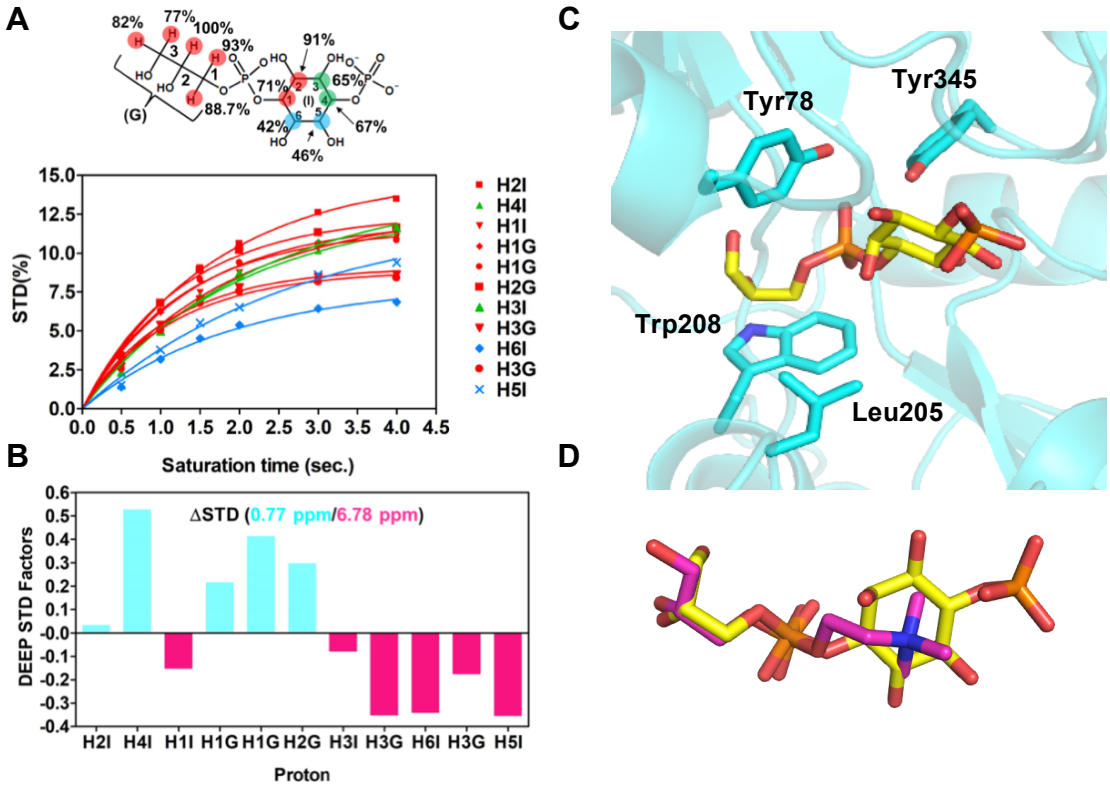


Figure 5.20: STD NMR of *Mtb* UgpB with GPI4P. A) Structure of glycerophosphoinositol-4-phosphate and experimental STD build up curve for the *Mtb* UgpB/GPIP4 complex and the obtained epitope map of GPI4P/*Mtb* UgpB. B) Differential epitope (DEEP)-STD factors showing the type of amino acid that the protons of the GPI4P ligand are orientated towards. Protons orientated towards aliphatic residues are highlighted in blue and protons orientated towards aromatic residues are highlighted in magenta. C) Docked structure of the GPIP4 in the binding site of *Mtb* UgpB. GPI4P is in stick representation with the carbon atoms in yellow. D) Close-up overlay of the binding orientations of GPC (magenta carbon atoms) with GPI(4)P (yellow carbon atoms). The figure was prepared by Dr Ridvan Nepravishta.

DEEP-STD NMR was also carried out to probe for additional information about how the GPI(4)P ligand is orientated in the binding site (Figure 5.20B) and it revealed that protons at positions 1 and 2 of the glycerol group (H1G), (H2G) and protons at position 4 of the inositol ring (H4I) are orientated towards the aliphatic residue Leu205 whereas protons at position 3 of the glycerol group (H3G) and positions 1, 3, 5 and 6 of the inositol ring (HI1), (H3I), (H5I) and (H6I) are orientated toward the aromatic Tyr78 and Tyr345 residues.

To investigate the 3D structural determinants of this interaction docking calculations were carried out using Autodock Vina to dock the experimentally obtained DEEP-STD factor values of GPI(4)P to the *Mtb* UgpB X-ray structure (Trott et al., 2010). By comparing the CORCEMA-ST calculated STD signals from the best scored docked structure of GPI(4)P bound *Mtb* UgpB and the experimental STD signals, a NOE R-factor of 0.31 was obtained and demonstrates very good data agreement of the docked structure to the experimental STD-NMR data as a NOE R-factor of 0.31-0.39 is typical for accurate structures (Edmondson, 1992) (Figure 5.20C). Interestingly alignment of the GPC ligand revealed that the glycerol tail of GPI(4)P had orientated differently possibly to accommodate the larger inositol-phosphate head-group (Figure 5.20D). Combined these data further suggest that the glyceorphosphodiester head-group size and charge are key molecular determinants for ligand selectivity.

5.4 Discussion

The structure of GPC bound *Mtb* UgpB was successfully solved by X-ray crystallography to 2.3 Å resolution. Initial attempts to co-crystallise GPC bound *Mtb* UgpB without chemical modification of the *Mtb* UgpB protein were unsuccessful. However determining the structures of dynamic and flexible proteins by X-ray crystallography is challenging as many molecules have to be aligned in the exact same orientation within the crystal lattice (Palamini et al., 2016). As SBP's undergo large conformational changes between the two-α/β domains, obtaining a ligand bound structure is difficult because the particular conformation needs to be stabilised, and this could explain the difficulty in obtaining a structure of D-glucosamine-6-phosphate bound *Mtb* UspC (Fullam et al., 2016). Large-scale motions may be necessary for ligand binding however previous methods to reduce flexibility and promote crystal formation include using a crystallisation chaperone for example an antibody (Iwata et al., 1995) or by controlled dehydration (Heras et al., 2003). Alternatively the surface of the protein can be targeted by mutating large, charged and surface exposed flexible amino acids e.g. lysine and glutamate to smaller hydrophobic residues (Lawson et al., 1991).

Instead of using those methods we chose to chemically modify the *Mtb* UgpB surface by reductive methylation of free amines e.g. lysine and the N-terminus amine. Reductive methylation of the surface exposed lysine residues of which there are potentially 13 modifiable residues likely improved the ability for GPC bound *Mtb* UgpB to crystallise as methylated lysines favour protein-protein interactions. In the GPC bound X-ray crystal structure only a single methylated lysine residue (MLZ 161) was identified. Structural alignment of GPC bound *Mtb* UgpB to the ligand unbound structure (PDB ID 4MFI) revealed as expected for SBP's that *Mtb* UgpB had undergone a large conformational change upon ligand binding associated with closure of the two-α/β domains around the GPC ligand. The conformational change was quantified by DynDom and CAVER analyses and indicated a rotation of 21.8° between domain I relative to domain II about the interdomain screw-axis which is comparable to other SBP's including MetQ that undergoes a 24° twist as a complex with MetNI (Nguyen et al., 2018). Compared to others SBP's this is moderate because the degree of rotation can be only a few degrees in the case of BtuF (Karpowich et al., 2003) and TroA (Y. H. Lee et al., 2002) or as large as 60° in the case of LivJ (Trakhanov et al., 2005).

In all of the *Mtb* UgpB molecules in the asymmetric unit, the GPC electron density was clearly defined and enabled the ligand and interacting amino acid residues to be accurately fitted. The residues involved in binding interacted primarily via the glycerol and phosphate moieties of GPC where Ser272, Ser153, Tyr78 and Tyr345 contacted the phosphate moiety *via* hydrogen bonding. Asp102 and Arg385 contacted the glycerol moiety also *via* hydrogen bonding. Trp208 contacted the glycerol backbone *via* a π-stacking interaction and Leu205 contacted the glycerol *via* Van Der Waals forces. Mutational analyses (Chapter 4, 4.3.13) supported that all of these residues were important for binding as replacement to alanine either completely abolished or reduced the GPC binding affinity as measured by microscale thermophoresis (MST). Interestingly no interactions occurred via the positively charged choline group despite the presence of this head-group being a key requirement for ligand binding by MST (Chapter 4, 4.3.12).

Compared to the G3P bound *E. coli* UgpB structure the binding pocket had similarly conserved residues however *Mtb* UgpB Leu205 and Asp102 specific residues were replaced by the larger aromatic Trp169 and the negatively charged Glu66 in *E. coli* UgpB (Wuttge et al., 2012). However when *Mtb* UgpB Leu205 was mutated to Trp169 (Jiang et al., 2014) no binding to GPC was observed by ITC. In *E. coli* UgpB Trp172, Tyr42 and Tyr323 also stabilise G3P and were conserved in *Mtb* UgpB. Similarly in *E. coli* UgpB Arg374 that binds to the C1 hydroxyl of G3P is conserved. As expected the ligand-binding pocket was more closed in G3P bound *E. coli* UgpB. Mutating *Mtb* UgpB Leu205 to alanine in a previous study (Jiang et al., 2014) also led to loss of binding to GPC supporting how Leu205 was essential for binding to GPC. These key differences in the structure of G3P bound *E. coli* UgpB compared to GPC bound *Mtb* UgpB likely reflect differences in the evolutionary origins of these organisms as *Mtb* is a well adapted as a human pathogen and may have evolved *Mtb* UgpB to scavenge glycerophosphodiesters as a nutrient source in macrophages.

In-solution methods were also used to investigate and support *Mtb* UgpB binding to GPC. STD-NMR experiments of the GPC *Mtb* UgpB interaction showed good agreement compared to the X-ray crystal structure. The ligand binding moieties were identified and indicated that H3G, H2G and H1G of the glycerol backbone showed the highest STD signals compared to weaker signals for H1C and H2C and the weakest signal for Me3C of the choline group. Furthermore DEEP STD-NMR analysis revealed that H3G of the glycerol group interacted with aliphatic residues modelled as Leu205 as it is the closest aliphatic residue in the ligand bound X-ray structure. H1C interacted

with aromatic residues and the closest residues identified in the crystal structure were Tyr78 and Tyr345 although these residues were actually identified to be interacting with the phosphate group. Saturation transfer difference NMR experiments also revealed that the alternative glycerophosphodiester GPI(4)P also binds to *Mtb* UgpB. The results indicated strong STD signals in regions of the inositol ring, part of the head-group whereas head-group interactions had not been previously observed for GPC suggesting that glycerophosphodiester size influences binding. Interestingly molecular docking based on STD-NMR measurements also found that the glycerol backbone of GPI(4)P re-orientated in order to accommodate the larger head-group supporting the idea that *Mtb* UgpB has the ability to accommodate and bind to alternative larger lipid head-groups.

Mtb UgpB was also co-crystallised with GPS and the putative GPS bound structure was solved by X-ray crystallography to 1.76 Å resolution. However although the electron density of the glycerol and phosphate groups was well defined, the serine group was not as well defined. It's possible that G3P is actually bound because the GPS ligand was produced to only 80 % purity. By NMR the impurity was glycerol however its possible that GPS is being hydrolysed leading to loss of the serine group during crystallisation conditions. However our microscale thermophoresis data (Chapter 4, 4.3.12) does not support binding to G3P. Interestingly the conformation of the putative GPS bound *Mtb* UgpB was almost identical to GPC bound *Mtb* UgpB. Also the orientation of the putative GPS ligand was identical to GPC, involving the glycerol and phosphate groups and no interactions with the head-group. Currently we cannot conclude that GPS is bound to *Mtb* UgpB based on this X-ray crystallography data therefore further studies are required.

This combined structural data suggests that *Mtb* UgpB is capable of binding to a wide range of glycerophosphodiester substrates and is possibly an evolutionary consequence of residing inside a nutrient limited environment. *Mtb* UgpB may be adapted to uptake glycerophosphodiesters derived from the degradation of membrane phospholipids generated by either bacterial or host derived phospholipases. It is interesting to note that GPC is a major component of lung surfactant and is also found at high levels in *Mtb* infected guinea pig granuloma (Somashekar et al., 2011). It's therefore possible that *Mtb* is in close proximity to large amounts of GPC that could be imported by the *Mtb* UgpABCE transporter enabling *Mtb* to survive in phosphate and carbon limited conditions.

Importantly our structural analysis of *Mtb* UgpB bound to GPC has facilitated the possibility of future medical chemistry research. Bacterial ABC importers provide access for essential nutrients therefore targeting importers is emerging as a new method to tackle antibiotic resistance (Tanaka et al., 2018). For example bacteriocins are natural antibiotics that mimic the natural substrates of several ABC transporters. Microcin C is a peptide-nucleotide antibiotic transported by the *E. coli yejABEF* ABC transporter and blocks $tRNA^{asp}$ (Novikova et al., 2007). Another example is albomycin that is an iron siderophore-antibiotic conjugate and is transported *via* the substrate binding protein of *E. coli* FhuD and inhibits seryl-tRNA synthetase (Clarke et al., 2002) (Rebuffat, 2012).

Synthetic antibiotics have also been designed to target and be imported by ABC transporters using a "Trojan horse like" method. For example the orthine carbamoyltransferase inhibitor, phaseolotoxin is imported by a dipeptide transporter *dppABCDF* (Pletzer et al., 2014) and kasugamcyin and blasticidin the translation initiation and termination inhibitors are transported by *E. coli oppABCDF* (Shiver et al., 2016). There are also reports of the SBP's of ABC transporters being directly targeted by drug candidates. For example RDS50 and RDS51 inhibited *S. typhimurium* growth and decreased invasion of intestinal epithelial cells (Ilari et al., 2016). Also X-ray crystallography revealed that the SBP interaction involved RDS51 co-ordinating a Zn^{2+} ion in the ZnuA SBP binding site, bridging the two-α/β domains of ZnuA by forming hydrophobic contacts and thereby either preventing ZnuABC transporter complex formation or the release of zinc (Ilari et al., 2016), resulting in starvation of essential Zn^{2+}.

Overall our structural analysis of *Mtb* UgpB provides a firm basis to design and develop inhibitors or substrate mimics. For example inhibitors could be designed based on the GPC bound *Mtb* UgpB structure, to contain a glycerol tail and phosphate group and a potentially toxic head-group moiety that could be used to target the *Mtb* UgpABCE transporter and elicit toxic effects inside the cell as antibiotics.

Chapter 6: Cloning, expression and purification of *Mycobacterium tuberculosis* UgpC and UgpAE

Now that we have biochemically characterised *Mtb* UgpB we wanted to understand the functions of the other components of the *Mtb* UgpABCE transporter therefore we decided to attempt to express and purify the nucleotide-binding protein *Mtb* UgpC (6A) and the membrane spanning proteins *Mtb* UgpAE (6B).

6A

6.1 Introduction

6.1.1 What do we know about *Mtb* UgpC

The *Mtb ugpC* gene (Rv2832c) encodes the predicted nucleotide-binding protein (NBP) of the *Mtb* UgpABCE transporter (Figure 6.1) and is annotated to be involved in energy coupling and active transport of *sn*-glycerol-3-phosphate across the plasma membrane based on genetic homology (Mycobrowser, 2019).

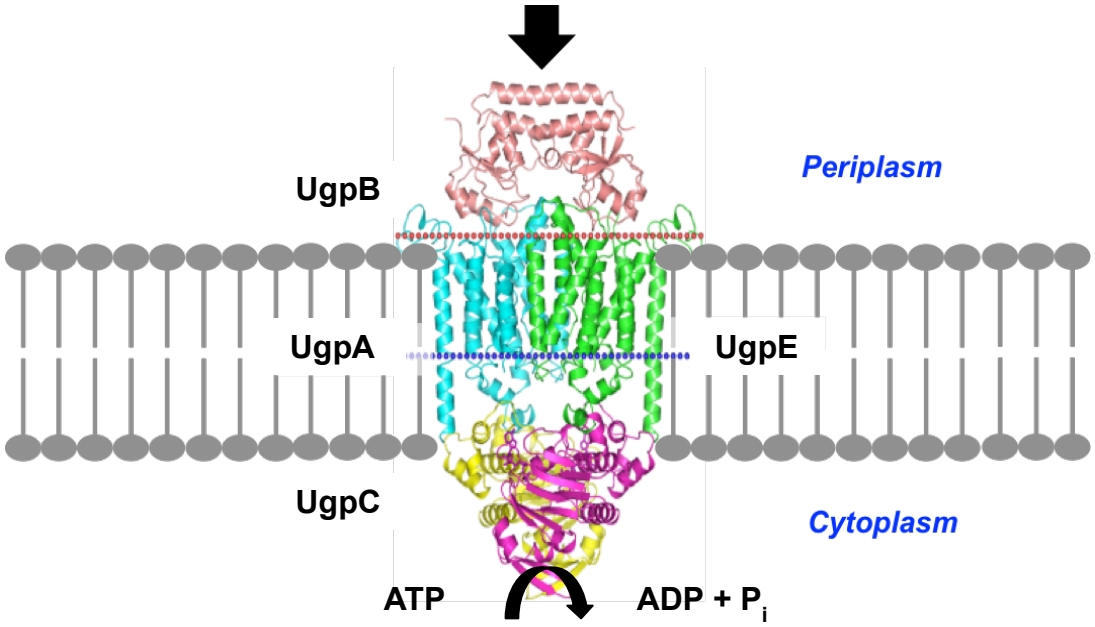


Figure 6.1: Predicted architecture of the *Mtb* UgpABCE ATP-binding cassette (ABC) transporter. UgpB is the substrate binding protein, UgpA and UgpE are the membrane spanning proteins and UgpC is the nucleotide binding protein.

The *Mtb ugpC* gene is located in the *Mtb ugpABCE* ATP-binding cassette (ABC) transporter operon (Chapter 3, 3.3.2) and is predicted to be the ATPase involved in coupling ATP hydrolysis and substrate import. Analysis of the *Mtb* UgpC amino acid

sequence by bioinformatics (Figure 6.2A) revealed that *Mtb* UgpC has the characteristic features of nucleotide-binding proteins (Locher, 2016) (Figure 6.2B). *Mtb* UgpC has a Walker-A motif shown in yellow (Figure 6.2A), predicted to bind to the α and β phosphates of ATP. A Walker-B motif shown in cyan (Figure 6.2A), that provides the catalytic glutamate residue, and a signature motif shown in green (Figure 6.2A), predicted to orientate ATP during hydrolysis.

A

MANVQYSAVTQRYPGADAPTVDNLDLDIADGEFLVLVGPSGCGKSTTLRVLAGLEPIE
SGRISIGDVDVTHLPPRARDVAMVFQNYALYPNMTVAANMGFALRNAGMSRADTRR
RVLEVADMLELTDLLDRKPAKLSGGQRQRVAMGRAIVRRPRVFCMDEPLSNLDAKLR
VSTRSQISGLQRRLGTTTVYVTHDQVEAMTMGDRVAVLKDGVLQQVDTPRALYDDP
VNTFVATFIGAPAMNLIDAAVAHGVVRAPDLAIPVPDPAAERVLVGVRPESWDVASIGT
PGSLTVHVELVEELGFESFVYATPVDQRGWSSRAPRIVFRTDRRTAVRVGESLAIVPH
SQEVRLFNSRTETRLR

B

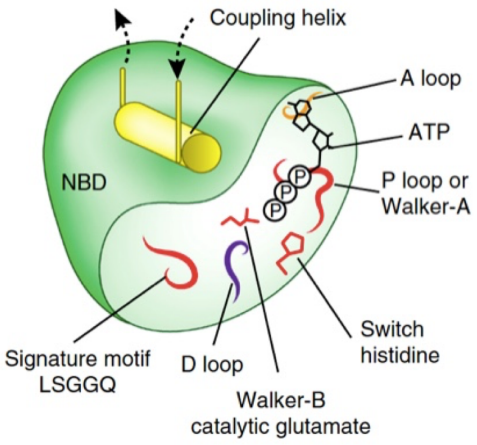


Figure 6.2: Amino acid sequence of *Mtb* UgpC and architecture of a single nucleotide-binding protein. A) Amino acid sequence of *Mtb* UgpC. The Walker-A motif (yellow), Walker-B motif (cyan) and signature motif (green) are highlighted. B) Architecture of an ABC-transporter single nucleotide binding protein showing features important for ATP binding and hydrolysis (Locher, 2016).

In order to biochemically investigate *Mtb* UgpC we wanted to clone, express, and purify *Mtb* UgpC and determine if it is a functional ATPase. Overall the aim is then to reconstitute the whole transporter and determine its structure, as this will help us to understand its function.

6.1.2 Assays to measure ATPase activity

ATPase activities can be characterised using several different methods. These include measuring phosphate release using colorimetric, fluorescent or radioactive substrates

(Lanzetta et al., 1979) (Brune et al., 1994) (Seals et al., 1978). A popular colorimetric assay for measuring phosphate release is the malachite green assay (Figure 6.3). The assay works by rapidly forming a coloured complex between malachite green, molybdate and phosphate in the presence of an acid. This coloured complex has an absorbance maximum at 600-660 nm and can be conveniently measured using a plate reader (Lanzetta et al., 1979). The advantages of the malachite green assay are that it is simple, very fast and highly sensitive, capable of measuring as little as 1.6 pmoles of phosphate. Therefore we decided to investigate this method.

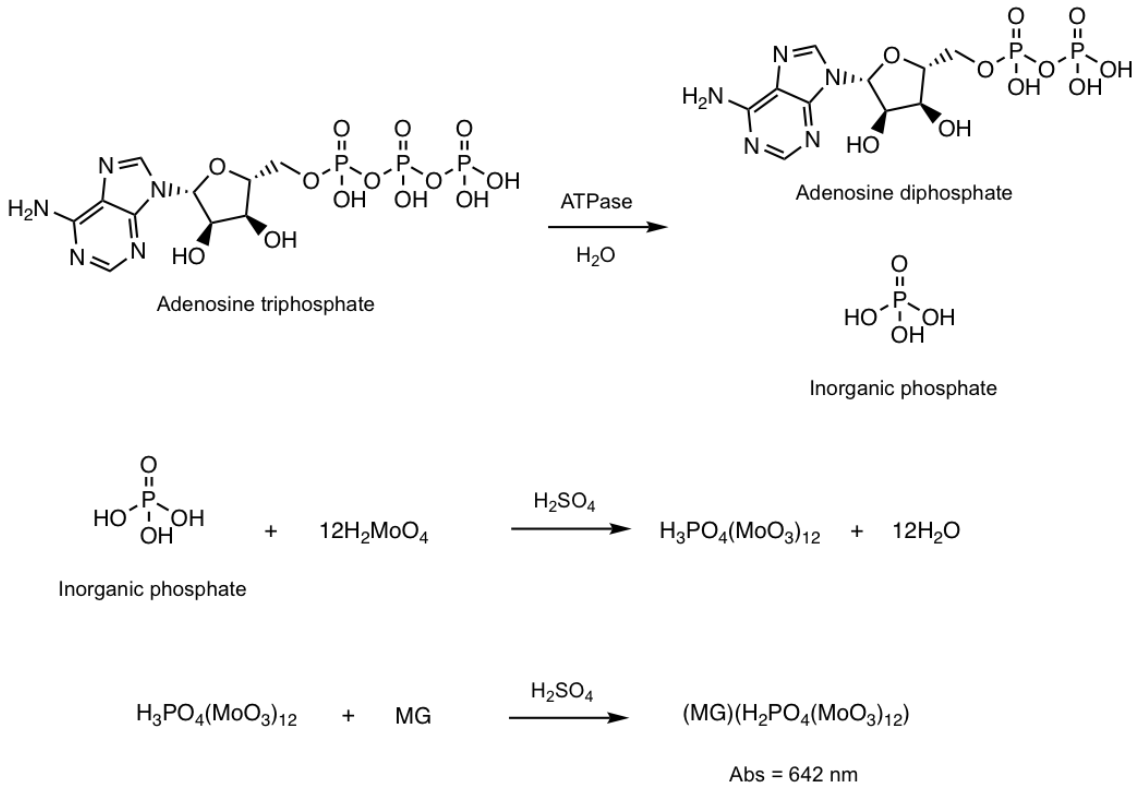


Figure 6.3: ATPase activity and the measurement of phosphate release by the malachite green assay. ATP is hydrolysed to ADP and P_i by ATPase activity. In the presence of acid, molybdate releases H_2O and forms a complex with free P_i. The resulting phosphomolybdate complex associates with malachite green (MG) producing a colour change that is measured by absorbance in the range of 600-660 nm.

Other methods to measure ATPase activity include using radioactive [γ-^{32}P] ATP substrates and measuring the release of radioactive $^{32}P_i$ (Seals et al., 1978). Fluorescent methods, for example the fluorescent reporter, *E. coli* phosphate binding protein (PBP) was conjugated to a fluorophore: N-[2-(1-maleimidyl)ethyl]-7-(diethylamino)coumarin-3-carboxamide (MDCC) and in the presence of phosphate forms a complex resulting in increased fluorescence emission (Brune et al., 1994). Other available fluorescent methods use enzyme cascades that convert phosphate into different substrates resulting in the release of hydrogen peroxide, detected by a specific fluorescent probe (CellBiolabs). Alternatively ADP release can be measured,

for example in the ADP-Glo^{TM} assay, ATPase activity releases ATP that is converted to ATP by a kinase enzyme, the ATP produced is then converted by a luciferase enzyme producing light (Zegzouti et al., 2009) (Promega).

6.2 Hypotheses and Aims

I hypothesised that the NBP *Mtb* UgpC could be expressed and purified in isolation without the membrane spanning subunits UgpA and UgpE. The protein could be purified to homogeneity using a combination of immobilised metal affinity and gel filtration chromatography, and its identity confirmed by mass spectrometry. In addition I hypothesised that the UgpC protein in the absence of the membrane spanning proteins, UgpA and UgpE would be a functional ATPase and I tested its activity using a malachite green ATPase assay.

The specific aims of the work presented in this chapter were:

- To clone, express and purify *Mtb* UgpC, the NBP of the UgpABCE ATP-binding cassette transporter and investigate if it is a functional ATPase.

6.3 Results

6.3.1 Cloning of *ugpC*-pET160Dest

ugpC-pET160Dest had been cloned previously in the lab using the Gateway cloning method. The vector (Figure 6.4) encodes the T7 IPTG inducible promoter and the full length *Mtb ugpC* (Rv2832c) gene with an N-terminal hexa-His tag. The vector also encodes the ampicillin resistance gene for selection purposes.

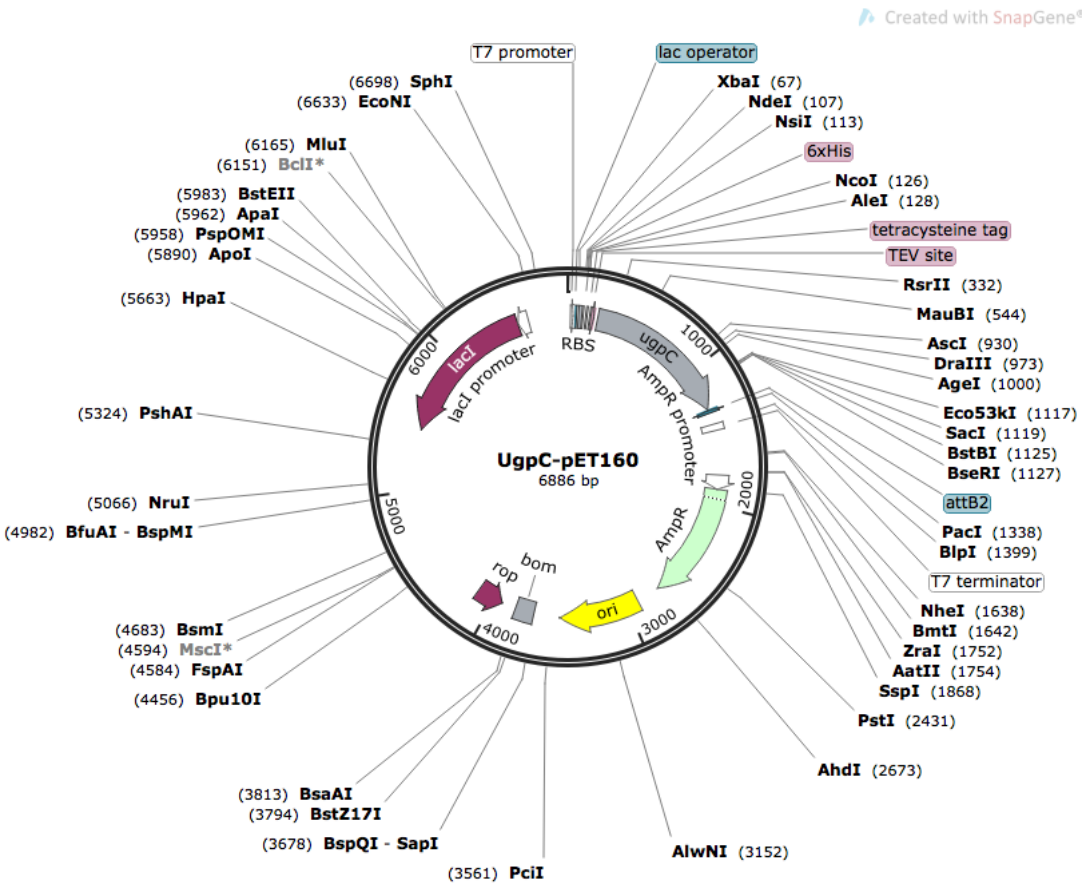


Figure 6.4: *Mtb ugpC*-pET160Dest plasmid map. Abbreviations: AmpR ampicillin resistance gene, ori origins of replication in *E. coli*.

6.3.2 Expression of UgpC-pET160Dest

The *ugpC*-pET160Dest construct was transformed into *E. coli* BL21 (DE3) competent cells, and incubated in 1 L of terrific broth containing ampicillin (100 µg/mL) with shaking at 180 rpm at 27 °C until optical density 0.6 was reached. Cells were induced by the addition of 1 mM IPTG and the culture temperature was lowered to 16 °C and expressed overnight. After expression the cells were suspended in UgpC buffer (300 mM NaCl, 20 mM tris-HCl, 20 % glycerol, 5 mM ATP, pH 7.5) (Chapter 2, Table 2.1) lysed by sonication and the cell lysate was passed over a Ni^{2+} immobilised metal affinity (IMAC) column. The bound proteins were eluted with a step-wise imidazole

gradient (Figure 6.5) however expression was not observed. *Mtb* UgpC was not produced as soluble protein as no band at 39 kDa was observed. It may have not been expressed or was insoluble. Therefore in order to optimise soluble expression we tried alternative *E. coli* strains and lower IPTG concentrations (1.0 and 0.1 mM) (6.3.3).

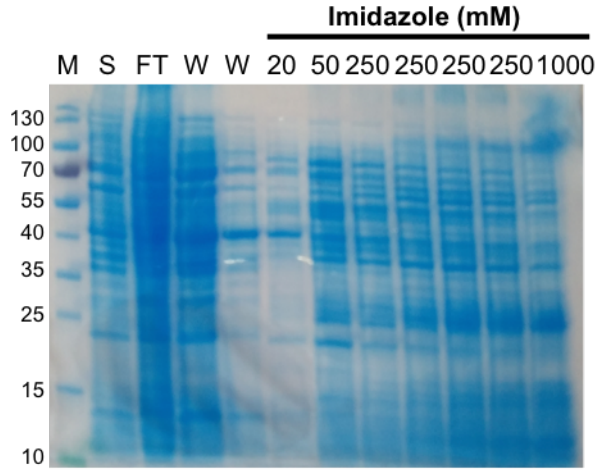


Figure 6.5: SDS-PAGE analysis of the expression of *Mtb* UgpC from *E. coli* BL21 (DE3). Elution of His_6-tagged *Mtb* UgpB from a Ni^{2+} IMAC-column. M = molecular weight marker in kDa, S = soluble lysate, W = washes, FT = flow through, numbers 20 – 1000 refer to the imidazole concentration in the elution buffer (units of mM).

6.3.3 Expression of UgpC-pET160Dest using different host strains

To improve expression of *Mtb* UgpC-pET160Dest it was also transformed into BL21 (DE3) pLysS cells, incubated in 100 mL of terrific broth culture media at 27 °C until optical density 0.6 was reached. The cells were induced by the addition of either 0.1 or 1 mM IPTG and expressed overnight at 16 °C. The cells after expression were suspended in UgpC buffer (300 mM NaCl, 20 mM tris-HCl, 20 % glycerol, 5 mM ATP, pH 7.5) (Chapter 2, Table 2.1) and lysed by sonication. The whole cell and soluble fractions were analysed by SDS-PAGE and Western blot (Chapter 2, 2.5.7) probed with a penta-His primary antibody and anti-mouse IgG alkaline phosphatase secondary antibody to investigate if the correct protein was expressed (Figure 6.6). *Mtb* UgpC expression could not be observed (Figure 6.6) at both IPTG concentrations or in *E. coli* BL21 (DE3) pLysS that control background protein expression through expression of T7 lysozyme. No expression was observed potentially because *Mtb* UgpC is membrane associated and previous studies have also found it is difficult to express NBP's (Nikaido et al., 1997).

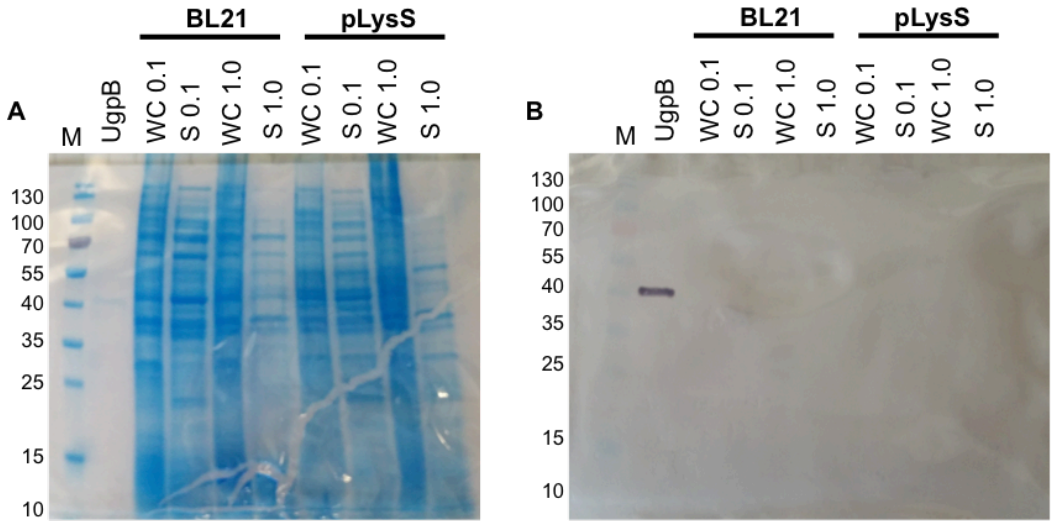


Figure 6.6: SDS-PAGE analysis and western blot of the expression of *Mtb* UgpC from *E. coli*. BL21 (DE3) and BL21 (DE3) pLysS A) SDS-PAGE analysis. M = molecular weight marker in kDa, UgpB = purified His_6-tagged *Mtb* UgpB, WC = whole cell extract, S = soluble lysate, numbers 0.1 and 1.0 refer to the IPTG concentration (units of mM), expression strains are labeled. B) Western blot of A probed with a penta-His primary antibody, anti-mouse IgG alkaline phosphatase secondary antibody and detected with BCIP alkaline phosphatase substrate.

6.3.4 Expression of UgpC-pET160Dest using different host strains and the addition of sorbitol:

Osmolytes such as sorbitol have been used to improve soluble protein expression when added at high concentrations to the culture media (Prasad et al., 2011). In this experiment *Mtb* UgpC-pET160Dest was transformed into BL21 (DE3), BL21 (DE3) pLysS and T7 Shuffle competent cells. T7 Shuffle cells promote protein disulphide bond formation in the cytoplasm. As there are two cysteine residues in *Mtb* UgpC that potentially form a disulphide bridge these cells may promote soluble protein expression. The cells were incubated in 100 mL of terrific broth culture media containing 0.5 M sorbitol until they reached optical density 0.6. The cells were then induced by the addition of either 0.1 or 1 mM IPTG and incubated overnight at 16 °C. After expression, the cells were suspended in UgpC buffer (Chapter 2, Table 2.1) lysed by sonication and the whole cell and soluble fractions were analysed by SDS-PAGE however *Mtb* UgpC expression could still not be observed (Figure 6.7). Instead of optimising further we chose to investigate using a different construct (6.3.5).

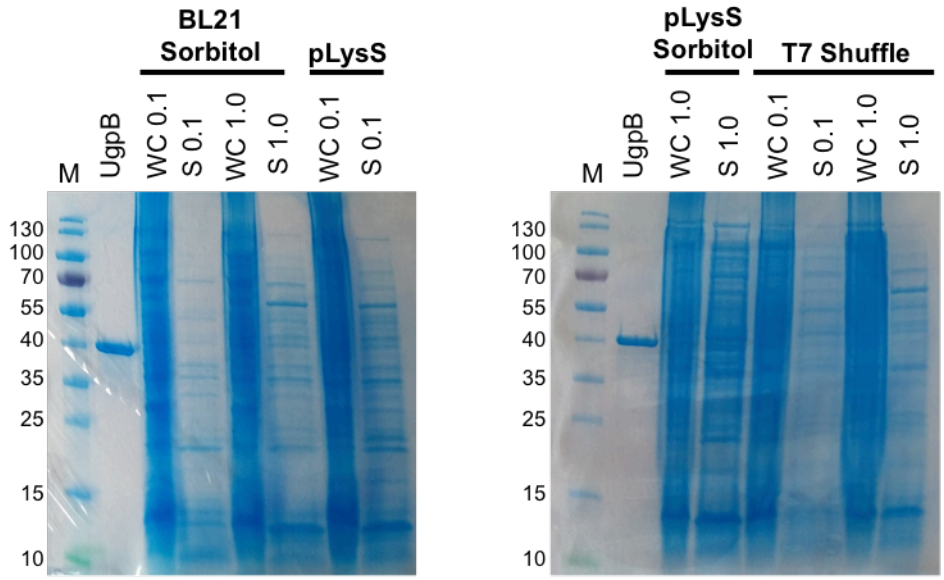


Figure 6.7: SDS-PAGE analysis of the expression of *Mtb* UgpC from *E. coli*. M = molecular weight marker in kDa, UgpB = purified His_6-tagged *Mtb* UgpB, WC = whole cell extract, S = soluble lysate, numbers 0.1 and 1.0 refer to the IPTG concentration (units of mM), expression strains are labeled, addition of sorbitol to the culture media is labeled.

6.3.5 Cloning of *ugpC*-pWaldo

Mtb ugpC-pWaldo was cloned previously in the lab. The plasmid (Figure 6.8) encodes the T7 IPTG inducible promoter and full length *Mtb ugpC* (Rv2832c) gene. The plasmid has a C-terminal TEV cleavable GFP tag for fluorescence detection. For purification purposes the plasmid has an octa-His tag. For selection purposes the plasmid has a kanamycin resistance gene. This was a potentially better expression vector because expression could be monitored by GFP fluorescence (Drew et al., 2001). Also by transforming the vector into *E. coli* Lemo21 cells the expression could be tuned using a L-rhamnose gradient to improve expression (Schlegel et al., 2012).

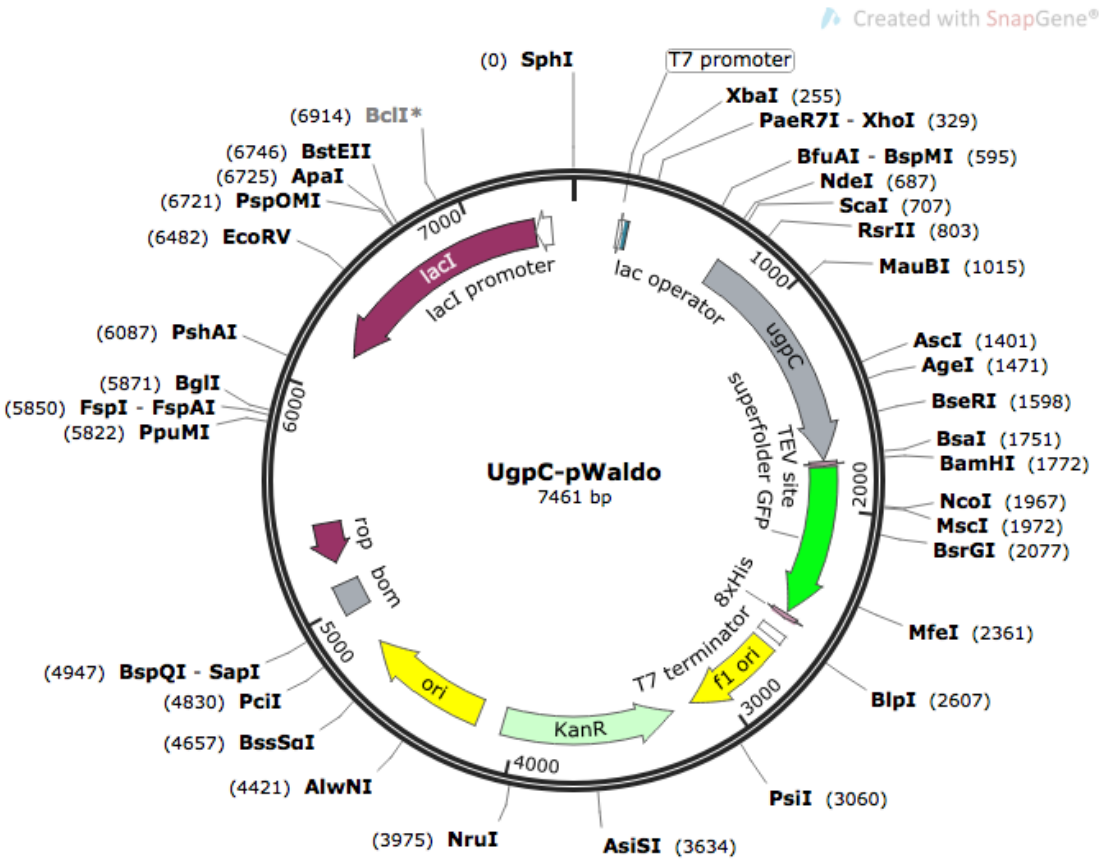


Figure 6.8: *Mtb ugpC*-pWaldo plasmid map. Abbreviations: KanR kanamycin resistance gene, f1 ori origin of replication in *E. coli*, GFP super folder green fluorescent protein.

6.3.6 Green fluorescent protein standard curve

To be able to calculate GFP expressed, a standard curve was produced (Figure 6.9) using purified GFP and measuring fluorescence in a plate reader at the following wavelengths, λ_{ex} 485 nm λ_{em} 535 nm (Tecan Infinite F200, gain 35). Purified GFP was used because the concentration was known and therefore fluorescence emission could be used to estimate the concentration of the GFP protein expressed. From this curve, 1 mg of GFP per L has 712 relative fluorescence units (RFU) at the measured gain of 35. For consistency a gain of 35 was used for all measurements. This standard curve was used to estimate the levels of *Mtb* UgpC-GFP protein expression in the L-rhamnose gradient experiment (6.3.7).

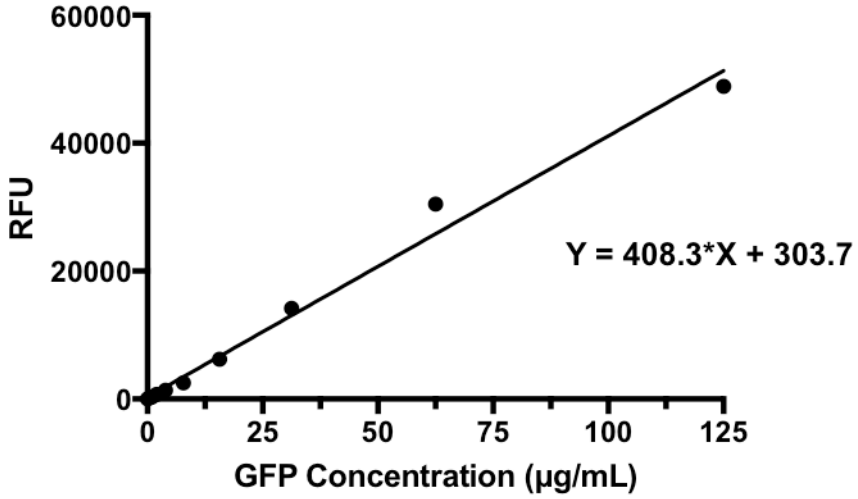


Figure 6.9: GFP standard curve. λ_{ex} 485 nm λ_{em} 535 nm (Tecan Infinite F200, gain 35).

6.3.7 Expression of UgpC-pWaldo in *E. coli* Lemo21 (DE3) cells

The UgpC-pWaldo construct was transformed into *E. coli* Lemo21 (DE3) cells and incubated in a small-scale trial in 5 mL of PASM-5052 media containing chloramphenicol (30 µg/mL) and kanamycin (50 µg/mL) (Chapter 2, Table 2.8) at 27 °C until optical density 0.6 was reached. At optical density 0.6, L-rhamnose was added to separate cultures to produce a L-rhamnose gradient from 0-1.0 mM (Table 6.1). The cultures were then either induced by addition of a single concentration of 0.4 mM IPTG or not induced, and incubated overnight at 16 °C. A single 0.4 mM IPTG concentration was used to evaluate the effect that different L-rhamnose concentrations had on protein expression. Following overnight expression GFP fluorescence of the whole cells was measured as described (Chapter 2, 2.5.3). The highest fluorescence reading of 39,294 RFU was observed for 0.1 mM L-rhamnose and 0.4 mM IPTG (Table 6.1). This fluorescence reading corresponded to approximately 9.5 mg of fluorescent protein per L of culture based on the GFP standard curve (Figure 6.9). As GFP fluorescence at the C-terminus normally indicates correct expressed protein folding this condition was scaled up in a 1 L expression experiment (6.3.8).

Table 6.1: L-rhamnose gradient expression of *Mtb* UgpC-GFP in *E. coli*. End-point optical density ($OD_{600\ nm}$) and fluorescence readings λ_{ex} 485 nm λ_{em} 535 nm.

IPTG (mM)	L-rhamnose (mM)	Optical density (OD_{600nm})	Fluorescence (RFU λ_{ex} 485 λ_{em} 535 nm)
0.4	0	0.177	19,129
0.4	**0.1**	**0.191**	**39,294**
0.4	0.25	0.207	32,739
0.4	0.5	0.194	9,729
0.4	0.75	0.194	4,816
0.4	1.0	0.201	5,235
0	0	0.201	12,327
0	0.1	0.214	20,694
0	0.25	0.194	25,097
0	0.5	0.199	12,797
0	0.75	0.190	6,402
0	1.0	0.194	12,052
Control Lemo	Control Lemo	0.195	80

Optical density readings are 1/20 dilution because 50 μL of cells were added to 950 μL PBS. Control Lemo refers to background fluorescence of *E. coli* Lemo21 cells that were not transformed with the construct. Red refers to the optimised expression condition.

6.3.8 1 L expression and purification of UgpC-pWaldo in *E. coli* Lemo21 (DE3) cells

Mtb UgpC-pWaldo *E. coli* Lemo21 was expressed in 1 L of PASM-5052 media containing chloramphenicol (30 μg/mL) and kanamycin (50 μg/mL) (Chapter 2, Table 2.8) using the optimised condition (6.3.7) (Table 6.1) and induced with 0.4 mM IPTG and 0.1 mM L-rhamnose. Following expression the cells were suspended in UgpC buffer (Chapter 2, Table 2.1) and lysed by sonication and purified by affinity of the C-terminal octa-His tag to a Ni^{2+} IMAC resin. Bound proteins were washed and eluted with a step-wise imidazole gradient and the fractions were analysed by SDS-PAGE (Figure 6.10). The expected *Mtb* UgpC-GFP fusion protein size was approximately 66 kDa. After SDS-PAGE, in-gel fluorescence showed a low fluorescence signal at 66 kDa and 26 kDa (Figure 6.10B). However staining with Instant Blue indicated no expressed band visible at 66 kDa (Figure 6.10A). This result was unchanged following TEV cleavage of the GFP tag (Figure 6.10C and 6.10D). Despite the high level of GFP fluorescence most of this signal was due to free GFP being expressed. This is known to be a problem if the protein is incorrectly folded (Bird et al., 2015).

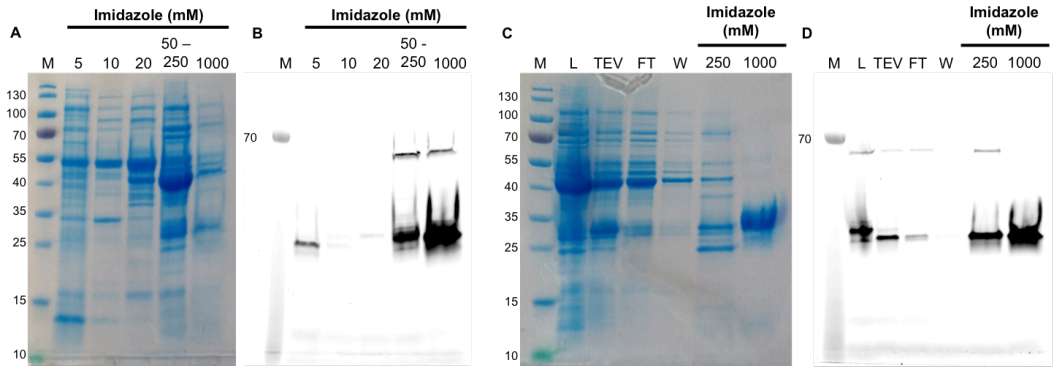


Figure 6.10: SDS-PAGE analysis and in-gel GFP fluorescence of the expression of *Mtb* UgpC-GFP from *E. coli* Lemo21 (DE3) cells. A) Elution of His_6-tagged *Mtb* UgpC-GFP from a Ni^{2+} IMAC-column. M = molecular weight marker in kDa, numbers 5 – 1000 refer to the imidazole concentration in the elution buffer (units of mM). B) In-gel GFP fluorescence of A, λ_{ex} 473 nm λ_{em} 502 nm. C) Elution of His_6-tagged TEV cleaved *Mtb* UgpC-GFP from a Ni^{2+} IMAC-column. M = molecular weight marker in kDa, L = protein before TEV cleavage, TEV = protein after TEV cleavage, FT = flow through, numbers 250 – 1000 refer to the imidazole concentration in the elution buffer (units of mM). D) In-gel GFP fluorescence of C, λ_{ex} 473 nm λ_{em} 502 nm.

6.3.9 Cloning of *ugpC*-pYUB1062 and *ugpC*-pYUB1062-GFP

A detailed procedure is described (Chapter 2, 2.3.3.2**).** *Mtb ugpC* was cloned into pYUB1062 and pYUB1062-GFP for expression in *M. smegmatis*. The difference between these constructs is that pYUB1062-GFP encodes a C-terminal GFP tag for monitoring protein expression.

Mtb ugpC was amplified from *Mtb* H37Rv genomic DNA using gene specific primers (Chapter 2, Table 2.2). Following amplification by PCR (Figure 6.11A and 6.12A), the restriction enzymes *NdeI* and *HindIII* were used to digest the PCR product and vectors at 37 °C for 3 hours. The digested products were then ligated with a 3:1 ratio of vector: insert DNA at 22 °C for 1 hour. The ligation products were transformed into *E. coli* Top10 and plated onto LB agar containing hygromycin (150 µg/mL) and incubated at 37 °C for 16 hours. Colonies that had formed after 16 hours were picked. PCR of each colony using vector specific primers (Chapter 2, Table 2.2) was used to determine if ligation had been successful. Six out of six colonies of *Mtb ugpC*-pYUB1062 *E. coli* Top10 screened were found to contain the *Mtb ugpC* insert (Figure 6.11B) and five out of five colonies of *Mtb ugpC*-pYUB1062-GFP *E. coli* Top10 screened were found to contain the *Mtb ugpC* insert (Figure 6.12B). The colonies were incubated in LB media containing hygromycin (150 µg/mL) and the plasmids were isolated (Chapter 2, 2.3.4). The plasmid from colony 3 (Figure 6.11B) and colony 4 (Figure 6.12B) were sequenced using vector specific primers (Chapter 2, Table 2.3) and found to have the correct *ugpC* insert sequences (Figure 6.13) (Figure 6.14).

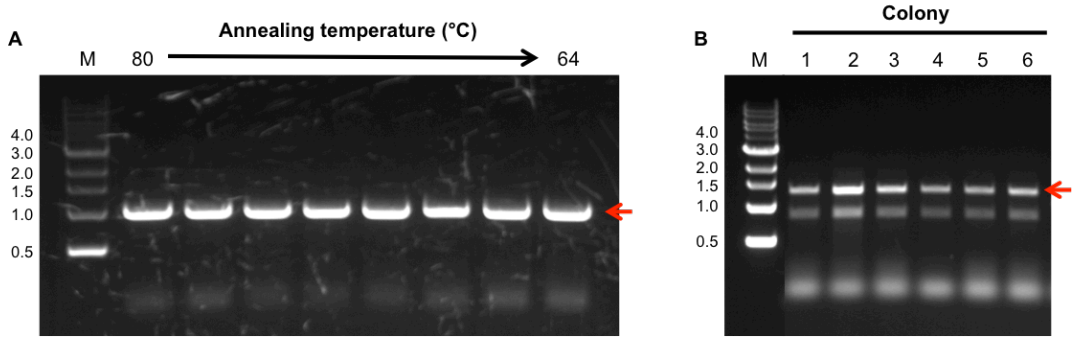


Figure 6.11: PCR amplification of *ugpC* and colony PCR of *ugpC*-pYUB1062. A) The *ugpC* gene was amplified from *Mtb* H37Rv genomic DNA using the primers 5'-aaaa*catatg*gctaacgttcagt-3' and 5'- aaaa*aagctt*gcgaagccgggtc-3'. The restriction enzyme sites are shown in italics. The PCR mixture contained: Q5 DNA polymerase (NEB), Q5 DNA polymerase buffer, 0.2 mM dNTP's, 6 % (v/v) DMSO and 0.5 µM forward and reverse primers. The PCR cycle consisted of: denaturation (98 °C, 30 s) followed by 30 cycles of annealing gradient (64-80 °C, 30 s), extension (72 °C, 90 s), followed by a final extension step (72 °C, 5 min). 20 µL of 50 µL PCR mixtures were loaded per lane onto a 1 % (v/v) agarose gel. M refers to 1 Kbp DNA ladder (Invitrogen). B) PCR of each colony using the specific primers (Chapter 2, 2.3.3.2) were used to determine whether the ligation had been successful.

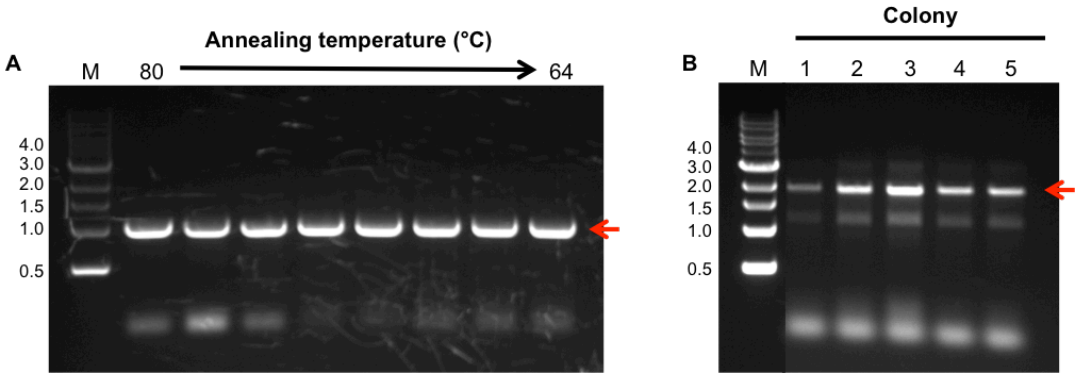


Figure 6.12: PCR amplification of *ugpC* and colony PCR of *ugpC*-pYUB1062-GFP. A) The *ugpC* gene was amplified from *Mtb* H37Rv genomic DNA using the primers 5'-aaaa*catatg*gctaacgttcagt-3' and 5'-aaaa*aagctt*tgcgaagccgggt-3'. The restriction enzyme sites are shown in italics. The PCR mixture contained: Q5 DNA polymerase (NEB), Q5 DNA polymerase buffer, 0.2 mM dNTP's, 6 % (v/v) DMSO and 0.5 µM forward and reverse primers. The PCR cycle consisted of: denaturation (98 °C, 30 s) followed by 30 cycles of annealing gradient (64-80 °C, 30 s), extension (72 °C, 90 s), followed by a final extension step (72 °C, 5 min). 20 µL of 50 µL PCR mixtures were loaded per lane onto a 1 % (v/v) agarose gel. M refers to 1 Kbp DNA ladder (Invitrogen). B) PCR of each colony using the specific primers (Chapter 2, 2.3.3.2) were used to determine whether the ligation had been successful.

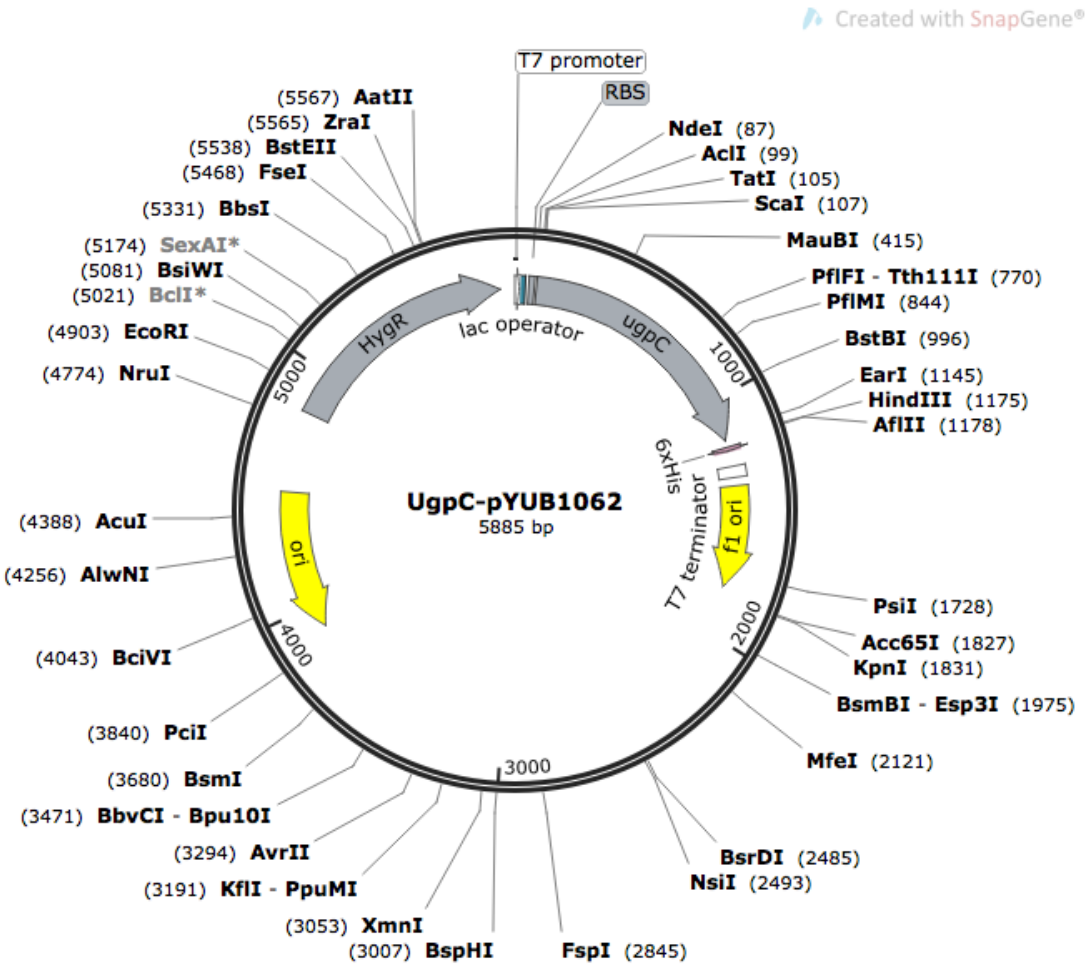


Figure 6.13: *Mtb* UgpC-pYUB1062 plasmid map. Abbreviations: HygR hygromycin resistance gene, f1 ori and f1 ori origins of replication in *E. coli* and *Mycobacteria*.

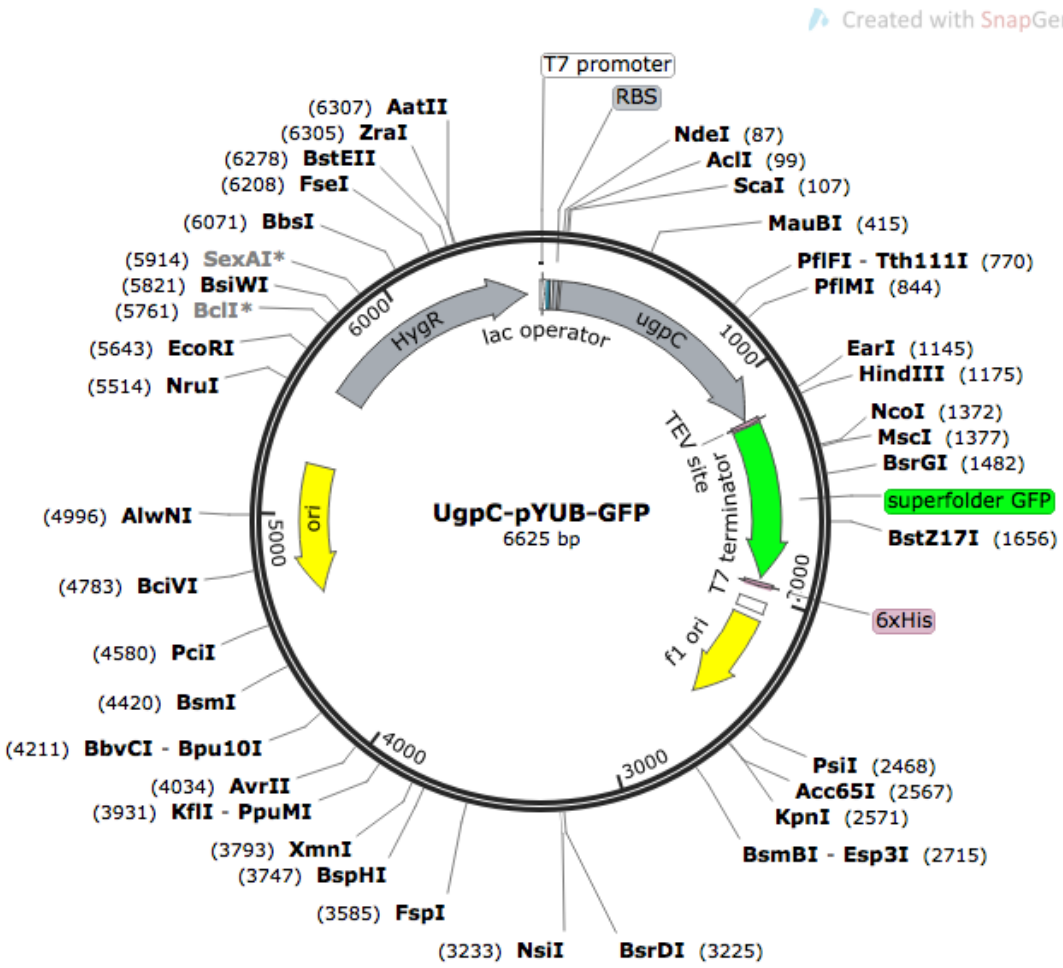


Figure 6.14: *Mtb* UgpC-pYUB1062-GFP plasmid map. Abbreviations: HygR hygromycin resistance gene, ori and f1 ori origins of replication in *E. coli* and *Mycobacteria*, GFP superfolder green fluorescent protein.

6.3.10 Expression and purification of UgpC-pYUB1062

The *ugpC*-pYUB1062 construct was electroporated into *M. smegmatis* $MC^2$4517 electro-competent cells (Chapter 2, 2.3.11) and incubated in 1 L of LBGT media (Chapter 2, Table 2.8) at 37 °C until optical density 0.6 was reached. The cells were induced by the addition of 0.2 % acetamide and further incubated overnight at 37 °C.

Following expression the cells were suspended in UgpC buffer (Chapter 2, Table 2.1), lysed by sonication and the expressed proteins were purified by affinity of the C-terminal hexa-histidine tag to a Ni^{2+} IMAC resin. Proteins that were bound to the IMAC column were washed and eluted with a step-wise imidazole gradient and fractions were analysed by SDS-PAGE (Figure 6.15). The expected *Mtb* UgpC protein size was approximately 39 kDa. After SDS-PAGE analysis there was no expressed band visible at 39 kDa (Figure 6.15). This result suggested that expressing *Mtb* UgpC in the more

closely related *M. smegmatis* host did not improve expression unlike the result observed for *Mtb* UgpB (Chapter 3, 3.3.4).

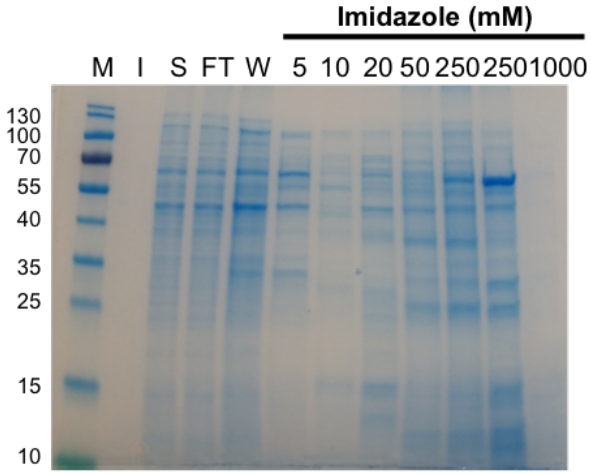


Figure 6.15: SDS-PAGE analysis of the expression of *Mtb* UgpC from *M. smegmatis*. Elution of His_6-tagged *Mtb* UgpC from a Ni^{2+} IMAC-column. M = molecular weight marker in kDa, I = insoluble fraction, S = soluble lysate, FT = flow through, W = wash, numbers 5 – 1000 refer to the imidazole concentration in the elution buffer (units of mM).

6.3.11 Expression and purification of UgpC-pYUB1062-GFP

The *ugpC*-pYUB1062-GFP construct was electroporated into *M. smegmatis* $MC^2$4517 cells (Chapter 2, 2.3.11), incubated in 1 L of LBGT media (Chapter 2, Table 2.8) at 37 °C until optical density 0.6 was reached. The cells were induced by the addition of 0.2 % acetamide and further incubated overnight at 37 °C. The GFP fluorescence of the whole cells (Chapter 2, 2.5.3) after expression was 21,194 RFU and corresponded to approximately 5.1 mg of protein per L of culture based on the GFP standard curve (Figure 6.9).

Following expression the cells were suspended in UgpC buffer (Chapter 2, Table 2.1), lysed by sonication and the expressed proteins were purified by affinity of the C-terminal hexa-histidine tag to a Ni^{2+} IMAC resin. Proteins that were bound to the IMAC column were washed and eluted with a step-wise imidazole gradient and fractions were analysed by SDS-PAGE (Figure 6.16). The expected *Mtb* UgpC-GFP fusion protein size was approximately 66 kDa. In-gel fluorescence (Figure 6.16B) revealed a low fluorescence signal at 66 kDa however a larger band was found at approximately 26 kDa that was likely to be free-GFP. This result was similar to UgpC-pWaldo expressed in *E. coli* Lemo21 cells. SDS-PAGE analysis revealed a band at approximately 66 kDa (Figure 6.16A). The 50-250 mM imidazole fractions containing the 66 kDa band were combined and passed over a Superdex 200 gel filtration chromatography column. However several contaminants co-eluted (Figure 6.16C) with the 66 kDa protein and it

was found to be low abundant (Figure 6.16D). This result confirmed that using an alternative more closely related *M. smegmatis* host to express *Mtb* UgpC was not successful under these conditions.

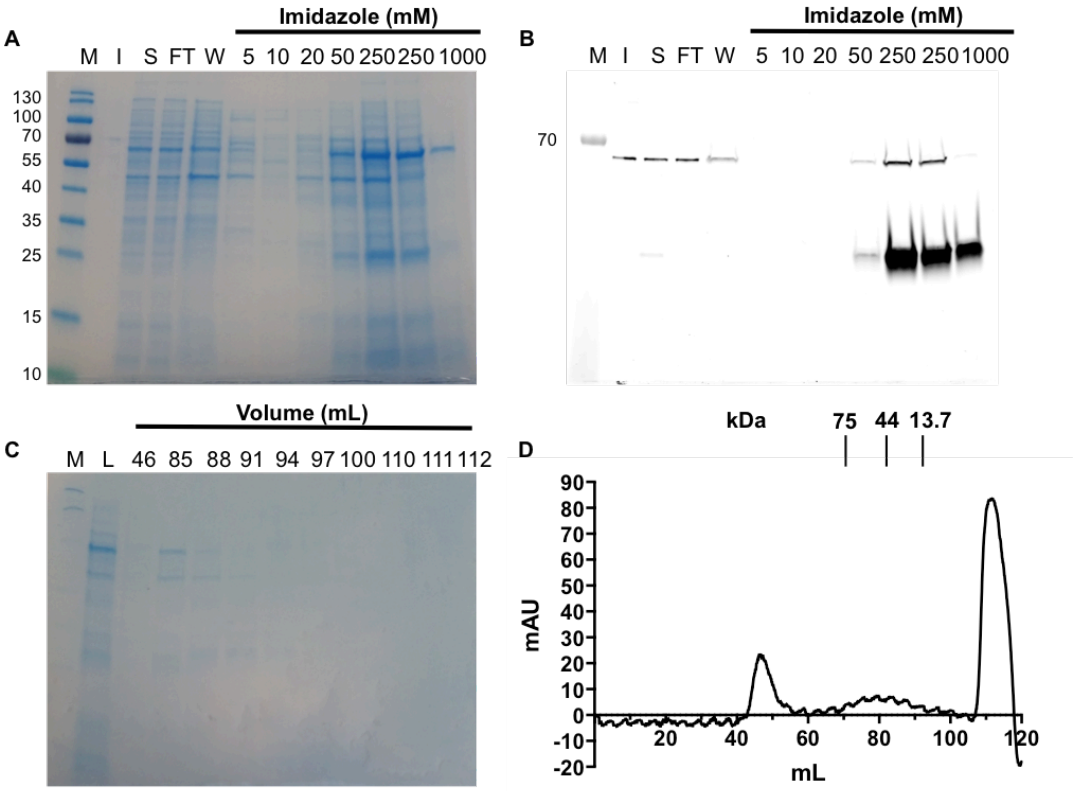


Figure 6.16: SDS-PAGE analysis and in-gel GFP fluorescence of the expression of *Mtb* UgpC-GFP from *M. smegmatis*. A) Elution of His_6-tagged *Mtb* UgpC-GFP from a Ni^{2+} IMAC-column. M = molecular weight marker in kDa, I = insoluble fraction, S = soluble lysate, FT = flow through, W = wash, numbers 5 – 1000 refer to the imidazole concentration in the elution buffer (units of mM). B). In-gel GFP fluorescence of A, λ_{ex} 473 nm λ_{em} 502 nm. C) Size exclusion chromatography of *Mtb* UgpC-GFP following a Ni^{2+} IMAC-column with the volumes shown as corresponding to D. L = protein before loading. D) Size exclusion trace of *Mtb* UgpC-GFP.

6.3.12 Cloning of *ugpC*-pETSUMO

Therefore the SUMO solubility tag was investigated, which has been successful for enhancing prokaryote protein expression (Peroutka Iii et al., 2011). A construct of *ugpC*-pETSUMO was not readily available therefore it was cloned. A detailed procedure is described (Chapter 2, 2.3.3.3**).**

Mtb ugpC was amplified from *Mtb* H37Rv genomic DNA using specific primers (Chapter 2, Table 2.2). Following amplification by PCR (Figure 6.17A), the restriction enzymes *BamHI* and *HindIII* were used to digest the PCR product and vector at 37 °C for 3 hours. The digested products were then ligated with a 3:1 ratio of vector: insert DNA at 22 °C for 1 hour. The ligation products were transformed into *E. coli* Top10 and

plated onto LB agar containing kanamycin (50 µg/mL) and incubated at 37 °C for 16 hours. Colonies that had formed after 16 hours were picked. PCR of each colony using specific primers (Chapter 2, Table 2.2) was used to determine if ligation had been successful. Six out of six colonies of *Mtb ugpC*-SUMO *E. coli* Top10 screened were found to contain the *Mtb ugpC*-SUMO insert (Figure 6.17B). The colonies were incubated in LB media containing kanamycin (50 µg/mL) and the plasmids were isolated (Chapter 2, 2.3.4). Sequenced of the plasmid from colony 2 confirmed it was correct (Figure 6.18).

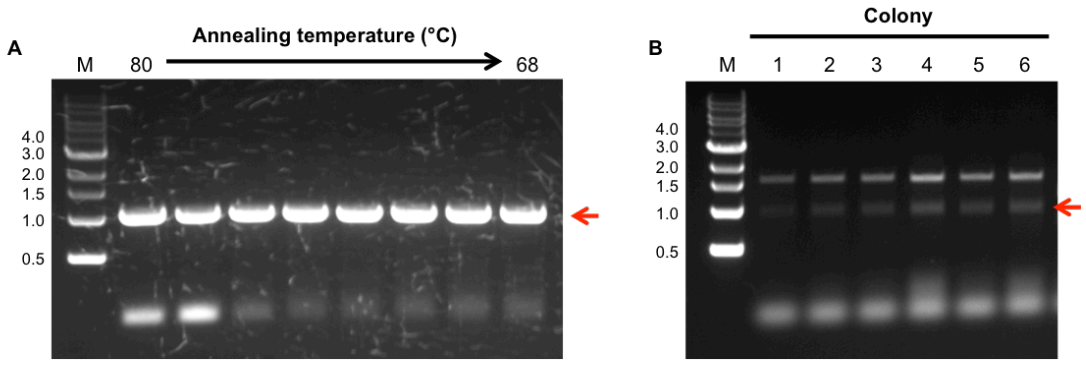


Figure 6.17: **PCR amplification of *ugpC* and colony PCR of *ugpC*-pETSUMO.** A) The *ugpC* gene was amplified from *Mtb* H37Rv genomic DNA using the primers 5'-aaaa*ggatcc*atggctaacgttc-3' and 5'-aaaa*aagctt*tcagcgaagccgg-3'. The restriction enzyme sites are shown in italics. The PCR mixture contained: Q5 DNA polymerase (NEB), Q5 DNA polymerase buffer, 0.2 mM dNTP's, 6 % (v/v) DMSO and 0.5 µM forward and reverse primers. The PCR cycle consisted of: denaturation (98 °C, 30 s) followed by 30 cycles of annealing gradient (68-80 °C, 30 s), extension (72 °C, 90 s), followed by a final extension step (72 °C, 5 min). 20 µL of 50 µL PCR mixtures was loaded per lane onto a 1 % (v/v) agarose gel. M refers to 1 Kbp DNA ladder (Invitrogen). B) PCR of each colony using the specific primers (Chapter 2, 2.3.3.3) were used to determine whether the ligation had been successful.

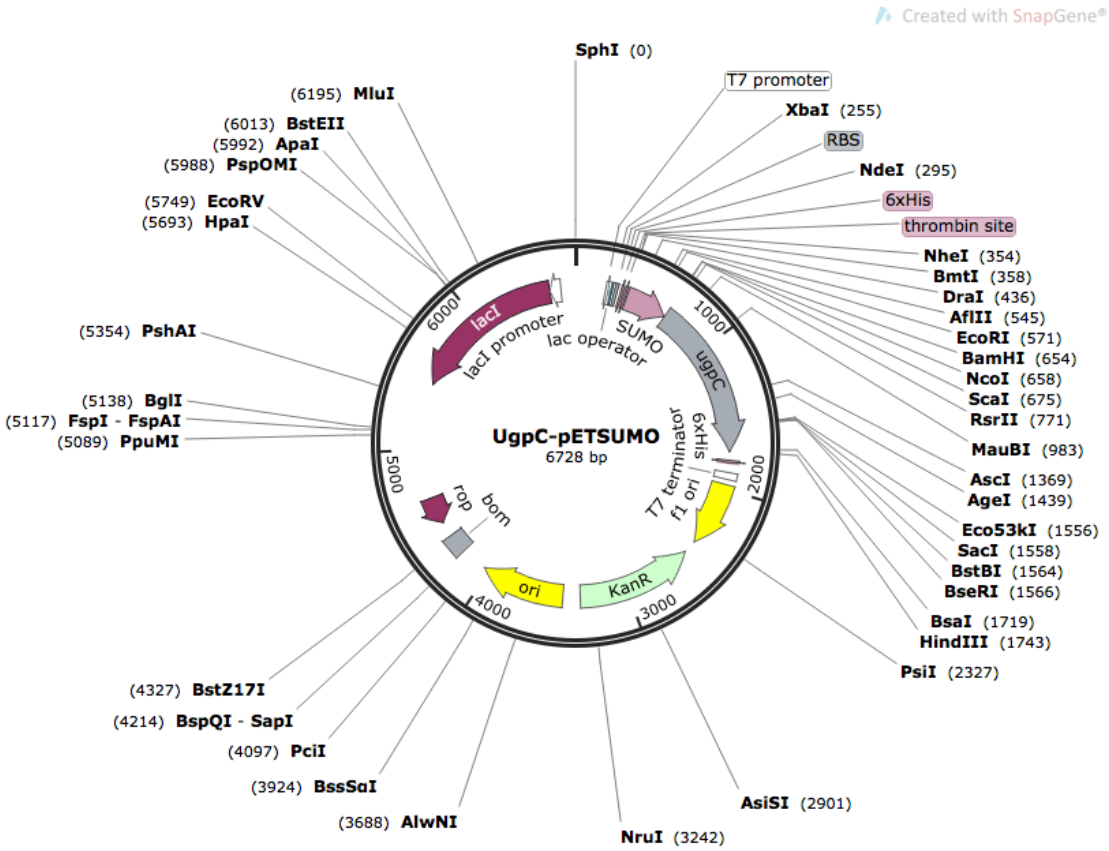


Figure 6.18: *Mtb ugpC*-pETSUMO plasmid map. Abbreviations: KanR kanamycin resistance gene, f1 ori origin of replication in *E. coli*, SUMO small ubiquitin-like modifier tag.

6.3.13 Expression and purification of UgpC-pETSUMO

The construct was transformed in *E. coli* BL21 (DE3) competent cells, incubated in 1 L of terrific broth at 27 °C until optical density 0.6 was reached. The cells were induced with 1 mM IPTG and further incubated overnight at 16 °C.

Following expression the cells were suspended in UgpC buffer (Chapter 2, Table 2.1) and lysed by sonication and the expressed proteins were purified by affinity of the N-terminal hexa-histidine tag to a Ni^{2+} IMAC resin. Bound proteins were washed and eluted with a step-wise imidazole gradient and fractions were analysed by SDS-PAGE (Figure 6.19). The expected *Mtb* UgpC-SUMO fusion protein size was approximately 50 kDa however only a small amount of the correct expressed protein was found in the gel (Figure 6.19A). The imidazole elution fractions (50-1000 mM) were combined and dialysed to remove the imidazole because imidazole can interfere with protease cleavage of the SUMO-tag. The dialysed fractions were then cleaved with a SUMO protease (2 mg/mL, 30 °C, 1 hour) and the digested protein was passed over a second Ni^{2+} IMAC column to remove the hexa-his SUMO tag (6.20B). A small amount of the *Mtb* UgpC protein was identified by the gel at approximately 39 kDa in the flow through

and wash step however a 60 kDa contaminant protein remained (Figure 6.19B). To increase the purity of the protein it was concentrated by ultrafiltration to 2 mL and passed over a Superdex 200 gel filtration chromatography column. However after gel filtration chromatography only a small amount of the 39 kDa protein was observed (Figure 6.19C). The protein was concentrated to 0.24 mg/ mL and stored at -80 °C prior to analyses.

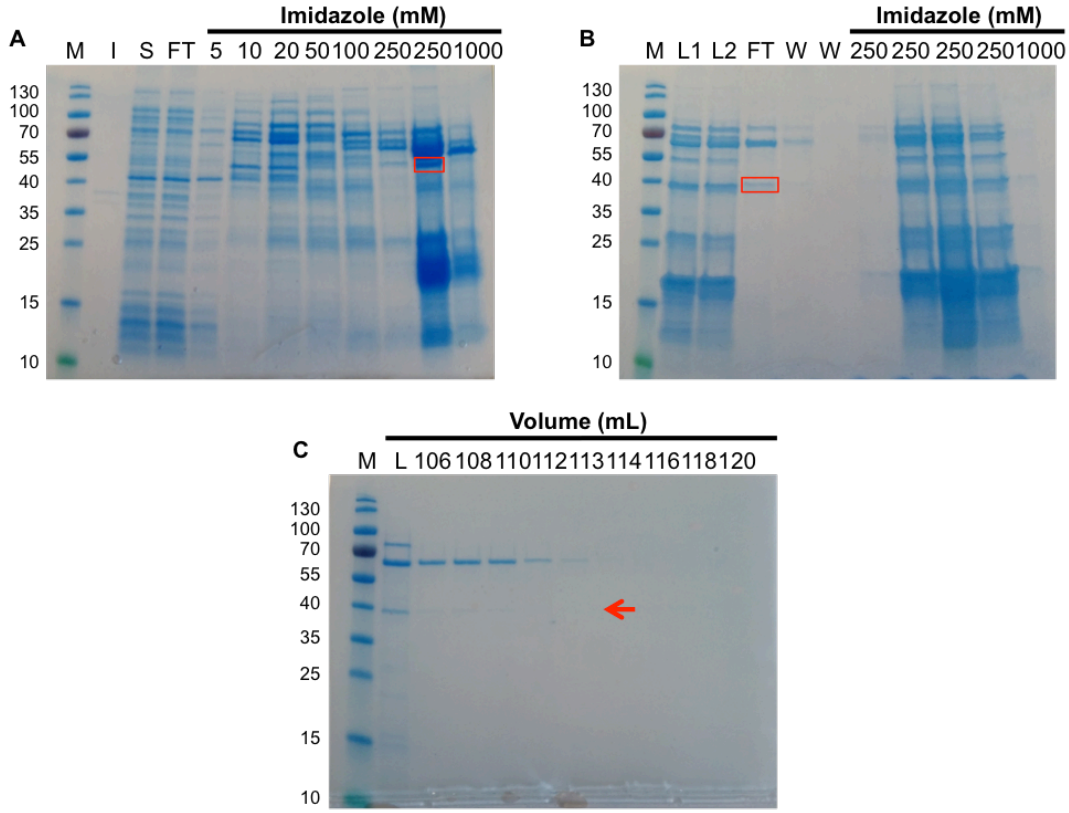


Figure 6.19: SDS-PAGE analysis of the purification of *Mtb* UgpC from *E. coli*. A) Elution of His_6-tagged *Mtb* UgpC from a Ni^{2+} IMAC-column. M = molecular weight marker in kDa, I = insoluble fraction, S = soluble lysate, FT = flow through, numbers 5 – 1000 refer to the imidazole concentration in the elution buffer (units of mM). B). Elution of His_6-tagged SUMO protease cleaved *Mtb* UgpC from a Ni^{2+} IMAC-column. M = molecular weight marker in kDa, L1 = protein after dialysis, L2 = protein after SUMO protease cleavage, FT = flow through, W = washes, numbers 250 – 1000 refer to the imidazole concentration in the elution buffer (units of mM). C) Size exclusion chromatography of *Mtb* UgpC following a Ni^{2+} IMAC-column with the volumes shown.

6.3.14 Mass spectrometry of UgpC-pETSUMO

The identity of the 39 kDa protein was successfully confirmed to be *Mtb* UgpC by in-gel trypsin digest mass spectrometry (Figure 6.20). 33 exclusive unique peptides, 61 exclusive unique spectra, 225 total spectra, 241/360 amino acids (67 % coverage).

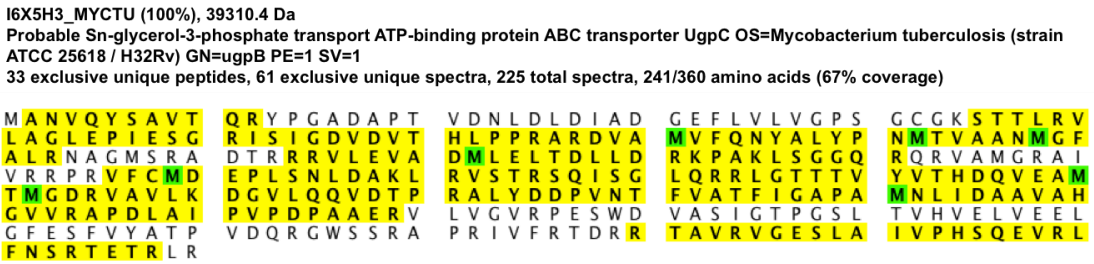


Figure 6.20: In-gel trypsin digest mass spectrometry of *Mtb* UgpC. Band extracted from figure 6.20C). Sequences highlighted in yellow indicate coverage of the peptides identified.

6.3.15 Large-scale expression of UgpC-pETSUMO

In order to increase the yield and purity of *Mtb* UgpC a larger 10 L culture was expressed using the same method (6.3.13).

6.3.16 Large-scale purification of UgpC-pETSUMO

The expressed proteins were purified as before (6.3.13). To increase the purity of the protein it was concentrated and passed over a Superdex 200 gel filtration chromatography column however several contaminant bands were still found (Figure 6.21C) therefore we decided to investigate the identity of these contaminants by mass spectrometry (6.3.17). The protein was concentrated to 1.05 mg/mL and stored at -80 °C prior to analyses.

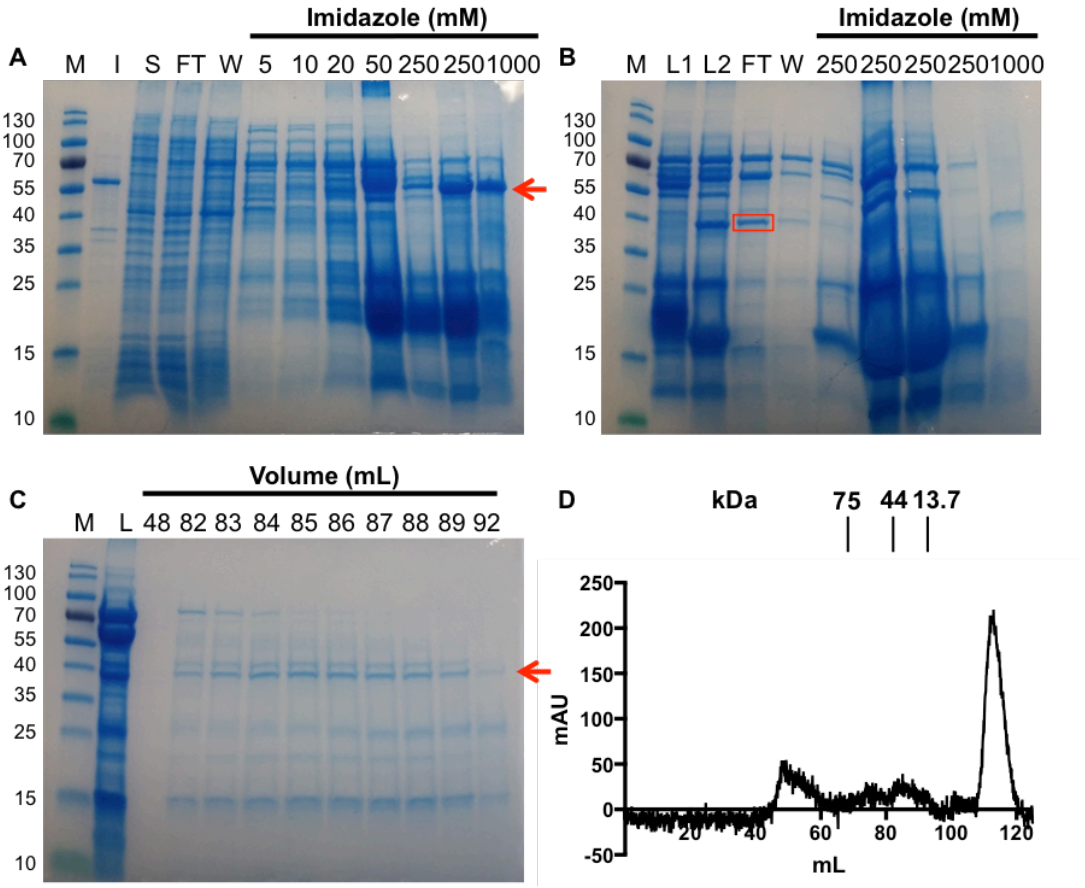


Figure 6.21: SDS-PAGE analysis of the purification of *Mtb* UgpC from *E. coli.* A) Elution of His_6-tagged *Mtb* UgpC from a Ni^{2+} IMAC-column. M = molecular weight marker in kDa, I = insoluble fraction, S = soluble lysate, FT = flow through, W = washes, numbers 5 – 1000 refer to the imidazole concentration in the elution buffer (units of mM). B). Elution of His_6-tagged SUMO protease cleaved *Mtb* UgpC from a Ni^{2+} IMAC-column. M = molecular weight marker in kDa, L1 = protein after dialysis, L2 = protein after SUMO protease cleavage, FT = flow through, W = washes, numbers 250 – 1000 refer to the imidazole concentration in the elution buffer (units of mM). C) Size exclusion chromatography of *Mtb* UgpC following a Ni^{2+} IMAC-column with the volumes shown as corresponding to D. D) Size exclusion trace of *Mtb* UgpC.

6.3.17 Mass spectrometry of UgpC-pETSUMO and contaminant proteins

In total 5 of the most abundant protein bands (Figure 6.21C) were analysed by in-gel trypsin digest mass spectrometry (Figure 6.22). The 69 kDa protein was identified to be a chaperone DnaK: 73 exclusive unique peptides, 143 exclusive unique spectra, 791 total spectra, 552/638 amino acids (87 % coverage). The 43 kDa protein was identified to be elongation factor tu tufA: 49 exclusive unique peptides, 109 exclusive unique spectra, 621 total spectra, 379/394 amino acids (96 % coverage). The 39 kDa protein was identified to be *Mtb* UgpC: 35 exclusive unique peptides, 108 exclusive unique spectra, 805 total spectra, 285/360 amino acids (79 % coverage). The 25 kDa protein was identified to be 30S ribosomal protein rpsC: 24 exclusive unique peptides, 37 exclusive unique spectra, 74 total spectra, 208/233 amino acids (89 % coverage). The 16 kDa protein was identified to be ferric uptake regulator protein fur: 23 exclusive

unique peptides, 51 exclusive unique spectra, 338 total spectra, 148/148 amino acids (100 % coverage). The identities of these proteins did not suggest they interacted with *Mtb* UgpC however they may have co-eluted due to the high concentration of glycerol and ATP present in the UgpC buffer.

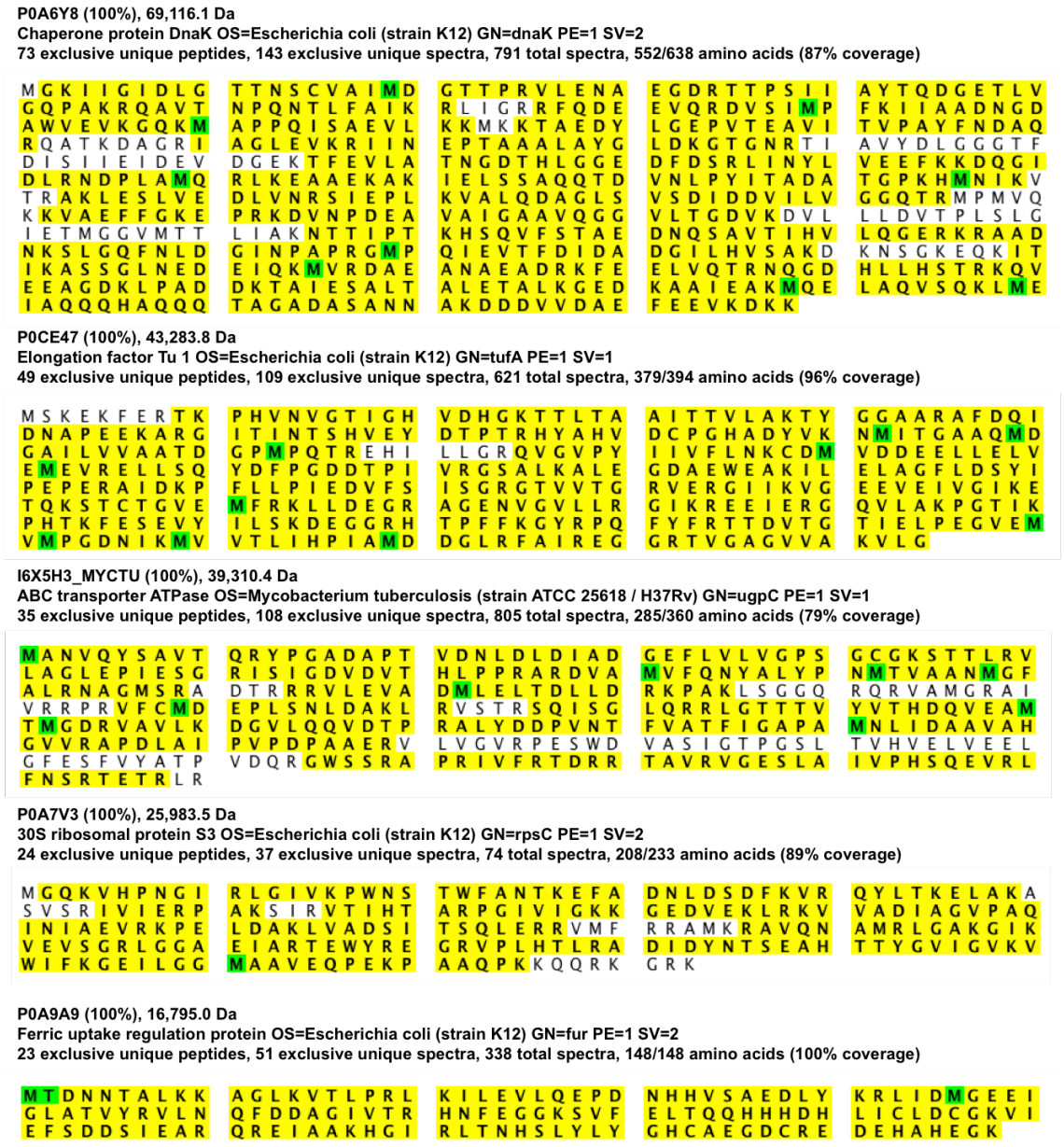


Figure 6.22: In-gel trypsin digest mass spectrometry of *Mtb* UgpC and contaminants. Bands extracted from Figure 6.21. Sequences highlighted in yellow indicate coverage of the peptides identified.

In conclusion the *Mtb* UgpC nucleotide binding protein of the UgpABCE transporter was successfully expressed and its identity was confirmed by mass spectrometry. Despite its purity being low it was used to investigate enzymatic activity (6.3.18)

6.3.18 UgpC malachite green phosphate assay

ATPase activity was investigated using a malachite green phosphate assay (Sigma Aldrich). In the assay, phosphate release from ATP hydrolysis results in the formation of a green colour complex between malachite green, molybdate and orthophosphate that can be measured by spectrophotometry (Figure 6.3). Firstly a standard curve was produced by dissolving the orthophosphate standard at different concentrations in assay buffer: 300 mM NaCl, 20 mM tris-HCl, 1 mM $MgCl_2$, 4 % glycerol (Figure 6.23). 20 μL of the malachite green assay reagent was added to 80 μL of orthophosphate standard and incubated at room temperature for 30 minutes prior to measuring absorbance at 595 nm as described (Chapter 2, 2.5.13).

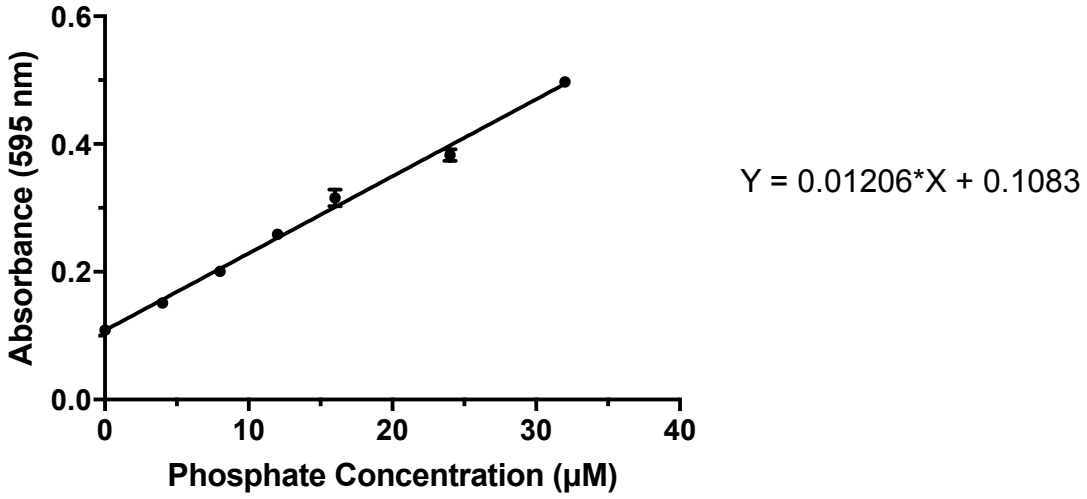


Figure 6.23: Malachite green phosphate standard curve.

The *Mtb* UgpC protein was dialysed into assay buffer (300 mM NaCl, 20 mM tris-HCl, 1 mM $MgCl_2$, 4 % glycerol) to remove the excess 5 mM ATP included throughout the purification steps and concentrated by ultrafiltration (10 kDa MWCO) to 0.6 mg/mL. In a 96-well plate, 80 μL of *Mtb* UgpC (0.6 mg/mL) and 0.1 mM ATP were added. Protein only and buffer only controls were also included. All of the reactions were incubated at 37 °C at 300 rpm for 3 hours. At selected time intervals, 20 μL of the malachite green assay reagent was added to stop the reactions and incubated at room temperature for 30 minutes to allow the colour to develop before measuring absorbance at 595 nm (Figure 6.24). The results indicated that *Mtb* UgpC + 0.1 mM ATP showed a greater absorbance compared to the protein only and buffer only controls, suggesting that ATP was being hydrolysed (Figure 6.24). However it's not clear why there was an increase in absorbance for the protein only control, it's possible that some ATP was still present

in the protein despite dialysis therefore the assay needs to be further optimised. As the protein preparation also contained *E. coli* DnaK, a chaperone known to have ATPase activity (Liberek et al., 1991) the purification needs to be optimised in order to confirm that *Mtb* UgpC is responsible for the observed ATPase activity. This assay was only carried out once. The assay would need to be repeated at least 3 times to show significance of the data.

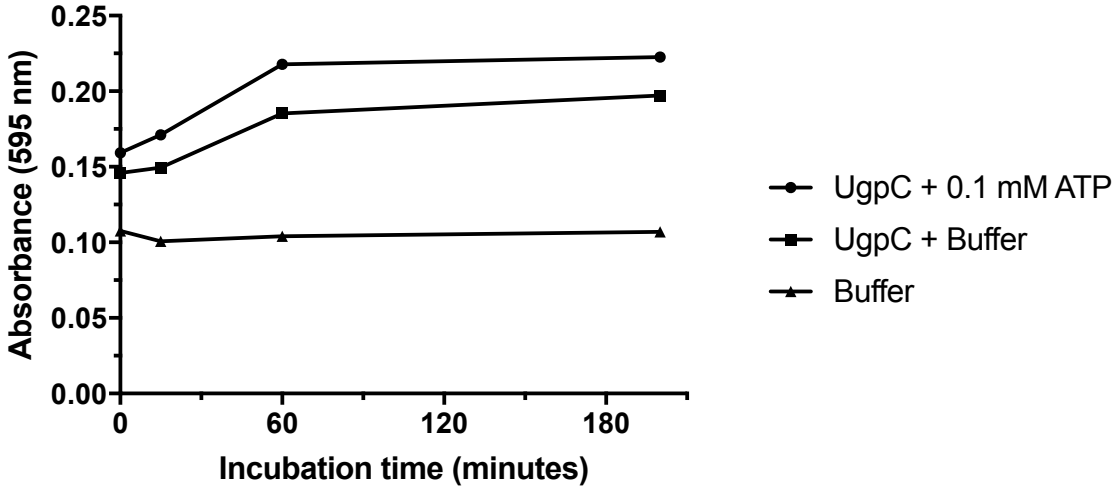


Figure 6.24: *Mtb* UgpC ATPase malachite green phosphate assay (n=1).

6B

6.4 Introduction

6.4.1 What do we know about *Mtb* UgpA and *Mtb* UgpE

We also wanted to express and purify *Mtb* UgpA (Rv2835c) and *Mtb* UgpE (Rv2834c), the two membrane spanning proteins of the UgpABCE transporter (Figure 6.1). This would facilitate crystal trials in order to determine the structure of *Mtb* UgpAE and aid in efforts to reconstitute the entire transporter. As elucidating the structure of *Mtb* UgpAE may reveal important features about the substrate transport mechanism.

Outside of the *Mycobacteria* genus the highest amino acid sequence similarity to *Mtb* UgpA and *Mtb* UgpE is the carbohydrate ABC transporter, membrane protein 1 of the CUT1 family from *Clotridioides difficile* (72.5 %), and the binding-protein-dependent transport system, inner membrane protein of *Clotridioides difficile* (69 %) respectively. This initially suggested that the *Mtb* UgpABCE transporter imports carbohydrates, however this is likely not the case because we found that the substrate binding protein, *Mtb* UgpB binds preferentially to glycerophosphodiesters (Chapter 4 and 5).

Topology mapping using Protter v1.0 (Omasits et al., 2014) predicted that *Mtb* UgpA and *Mtb* UgpE are both integral membrane proteins. Each is made up of 6 transmembrane helices and both the N- and C- termini are found intracellularly (Figure 6.25). As integral membrane proteins, *Mtb* UgpA and *Mtb* UgpE are anticipated to be difficult to express and purify (Schlegel et al., 2014). Therefore we chose to clone *Mtb ugpAE* in the pWaldo vector and express in *E. coli* Lemo21 (DE3) cells as this system has previously been used to successfully express membrane proteins (Drew et al., 2001).

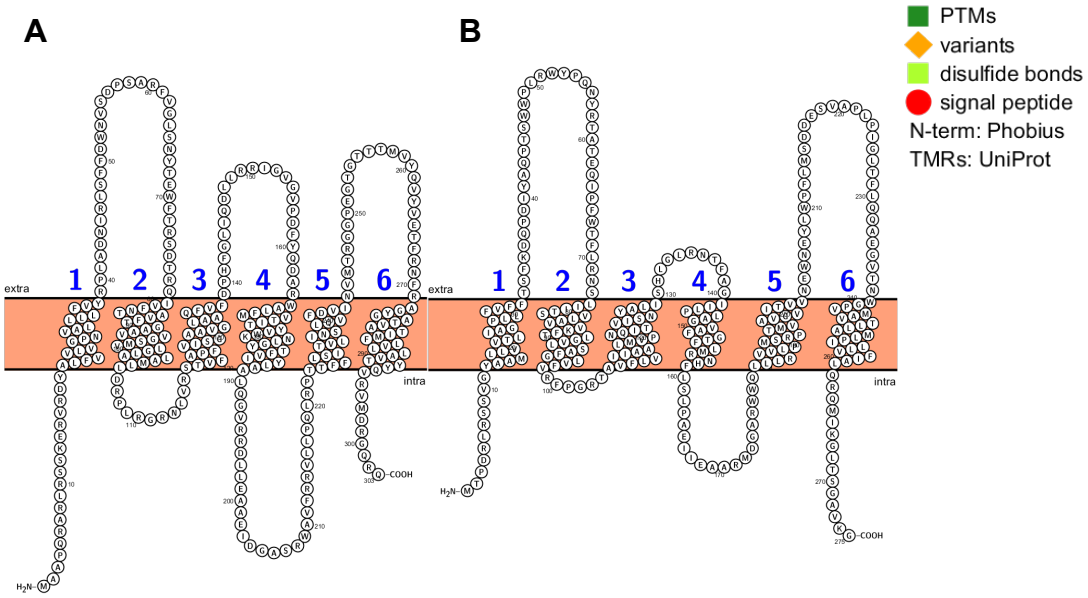


Figure 6.25: Predicted topology of *Mtb* UgpAE mapped using Protter v1.0 (Omasits et al., 2014). A) *Mtb* UgpA (Uniprot accession number: I6XFF3). B) *Mtb* UgpE (Uniprot accession number: A5U6I6)

6.5 Hypotheses and Aims

I hypothesised that the membrane spanning proteins, UgpA and UgpE could be expressed using the same origin of replication in the pWaldo vector using an L-rhamnose tuneable *E. coli* cell expression system. I tested this hypothesis by measuring GFP fluorescence to report protein expression.

The specific aims of the work presented in this chapter were:

- To clone, express and purify *Mtb* UgpAE, the membrane-spanning proteins of the UgpABCE ATP-binding cassette transporter.

6.6 Results

6.6.1 Cloning of *ugpAE*-pWaldo

A detailed cloning procedure is described (Chapter 2, 2.3.3.4). *Mtb ugpAE* was amplified from *Mtb* H37Rv genomic DNA using specific primers (Chapter 2, Table 2.2). Following amplification by PCR (Figure 6.26A), the restriction enzymes *NdeI* and *BamHI* were used to digest the PCR product and vector at 37 °C for 3 hours. The digested products were then ligated with a 3:1 ratio of vector: insert DNA at 22 °C for 1 hour. The ligation products were transformed into *E. coli* Top10 and plated onto LB agar containing kanamycin (50 µg/mL) and incubated at 37 °C for 16 hours. Colonies that had formed after 16 hours were picked. PCR of each colony using specific primers (Chapter 2, Table 2.2) was used to determine if ligation had been successful. Six out of six colonies of *Mtb* UgpAE-pWaldo *E. coli* Top10 screened were found to contain the *Mtb* UgpAE insert. The colonies were grown in LB media containing kanamycin (50 µg/mL) and the plasmids were isolated (Chapter 2, 2.3.4). The plasmid from colony 5 was found to have the correct sequence (Figure 6.27).

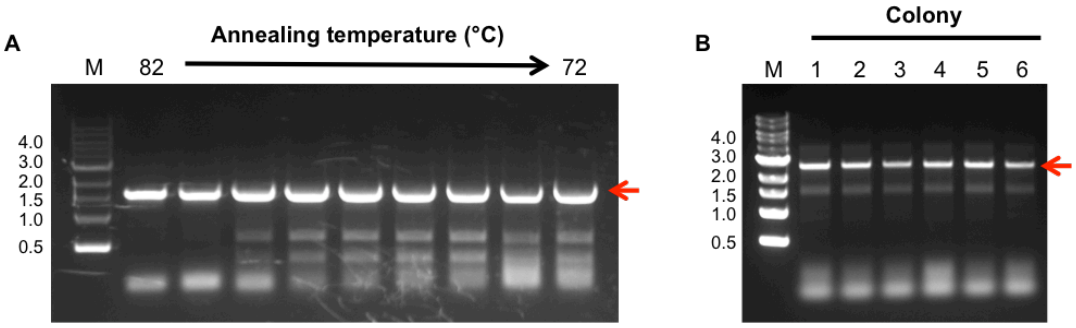


Figure 6.26: PCR amplification of *ugpAE* and colony PCR of *ugpAE*-pWaldo. A) The *ugpAE* gene was amplified from *Mtb* H37Rv genomic DNA using the primers 5'-aaaaaa*catatg*gcggcgccgca-3' and 5'-ttttttt*ggatcc*tcccttgaccg-3'. The restriction enzyme sites are shown in italics. The PCR mixture contained: Q5 DNA polymerase (NEB), Q5 DNA polymerase buffer, 0.2 mM dNTP's, 6 % (v/v) DMSO and 0.5 µM forward and reverse primers. The PCR cycle consisted of: denaturation (98 °C, 30 s) followed by 30 cycles of annealing gradient (72-82 °C, 30 s), extension (72 °C, 90 s), followed by a final extension step (72 °C, 5 min). 20 µL of 50 µL PCR mixtures was loaded per lane onto a 1 % (v/v) agarose gel. M refers to 1 Kbp DNA ladder (Invitrogen). B) PCR of each colony using the specific primers (Chapter 2, 2.3.3.4) were used to determine whether the ligation had been successful.

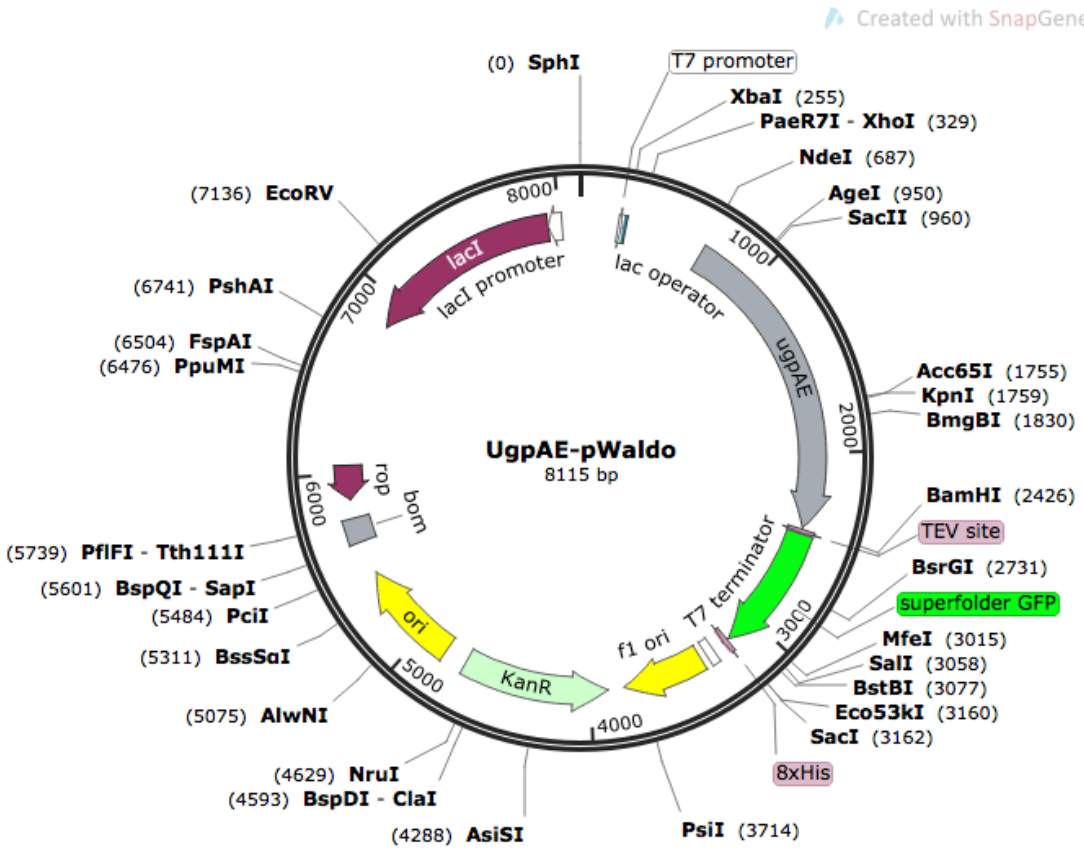


Figure 6.27: *Mtb* UgpAE-pWaldo plasmid map. Abbreviations: KanR kanamycin resistance gene, f1 ori origin of replication in *E. coli*, GFP super folder green fluorescent protein.

6.6.2 Expression and purification of UgpAE-pWaldo

The *ugpAE*-pWaldo construct was transformed into *E. coli* Lemo21 (DE3) cells. The C-terminal GFP tag was used to monitor expression. The cells were incubated in a small-scale 5 mL PASM-5052 media containing chloramphenicol (30 µg/mL) and kanamycin (50 µg/mL) (Chapter 2, Table 2.8) at 27 °C until optical density 0.6 was reached. At optical density 0.6, L-rhamnose was added to separate cultures to produce a L-rhamnose gradient from 0-1.0 mM (Table 6.2). The cultures were then either induced by addition of a single concentration of 0.4 mM IPTG or not induced, and incubated overnight at 16 °C. A single 0.4 mM IPTG concentration was used to evaluate the effect that different L-rhamnose concentrations had on protein expression. Following overnight expression GFP fluorescence of the whole cells was measured as described (Chapter 2, 2.5.3). The fluorescence was found to be low in all the conditions tested (Table 6.2). To investigate protein expression the cells from conditions highlighted in red (Table 6.2) were suspended in UgpAE purification buffer (Chapter 2, Table 2.1), lysed by sonication and the whole cell and soluble fractions were analysed by in-gel fluorescence and SDS-PAGE (Figure 6.28). The approximate 90 kDa UgpAE-GFP protein could not be identified indicating that expression was not successful. This could

be due to the toxic effects of expressing an integral membrane protein in *E. coli.* Due to time, further optimisation was not possible.

Table 6.2: L-rhamnose gradient expression of *Mtb* UgpAE-GFP in *E. coli*. End-point optical density ($OD_{600\ nm}$) and fluorescence readings λ_{ex} 485 nm λ_{em} 535 nm.

IPTG (mM)	L-rhamnose (mM)	Optical density (OD_{600nm})	Fluorescence (RFU λ_{ex} 485 λ_{em} 535 nm)
0.4	0	0.182	251
0.4	0.1	0.196	103
0.4	0.25	0.205	94
0.4	0.5	0.189	90
0.4	0.75	0.192	79
0.4	1.0	0.182	70
0	0	0.193	223
0	0.1	0.18	168
0	0.25	0.181	149
0	0.5	0.187	124
0	0.75	0.195	103
0	1.0	0.182	95
Control Lemo	Control Lemo	0.187	80

Optical density readings are 1/20 dilution because 50 μL of cells were added to 950 μL PBS. Control Lemo refers to background fluorescence of *E. coli* Lemo21 cells that were not transformed with the construct. Red refers to the optimised expression condition.

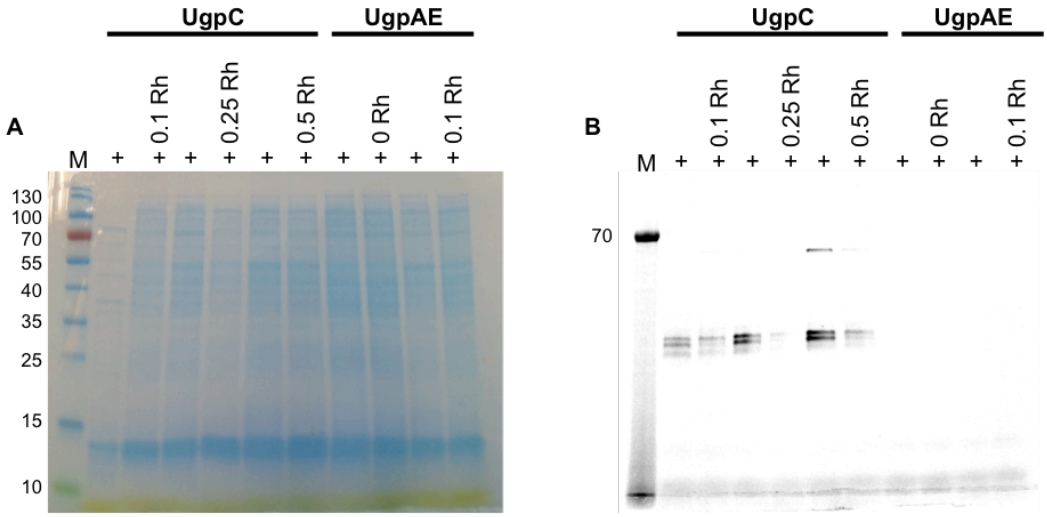


Figure 6.28: SDS-PAGE analysis and in-gel GFP fluorescence of the expression of *Mtb* UgpC and *Mtb* UgpAE from *E. coli*. A) SDS-PAGE analysis of whole cell extracts. M = molecular weight marker in kDa, + = 0.4 mM IPTG, Rh = L-rhamnose concentration (mM). B) In-gel GFP fluorescence of A , λ_{ex} 473 nm λ_{em} 502 nm.

6.7 Discussion

The aim of this chapter was to express and purify the NBP *Mtb* UgpC, and to investigate if it is a functional ATPase. An additional aim was to express and purify the membrane spanning proteins *Mtb* UgpAE. To attempt to express *Mtb* UgpC a range of different conditions were tested, these included: using different host strains including *E. coli* BL21 (DE3), pLysS, T7 Shuffle and the more closely related *M. smegmatis* host. Different IPTG concentrations were used and the osmolyte, sorbitol was added to the media. Different expression systems including a L-rhamnose tuneable expression system was used. GFP tags were used to monitor protein expression and the SUMO solubility enhancement tag was also used.

From these studies we found we were able to successfully express *Mtb* UgpC with an N-terminal SUMO tag in *E. coli* (6.3.16), which was confirmed by mass spectrometry (6.3.17). However disappointingly, purification of *Mtb* UgpC proved to be difficult as several contaminants co-eluted following gel filtration chromatography. Previous studies have also had difficulties expressing NBP's. For example HisP, the histidine permease NBP from *Salmonella typhimurium* (Nikaido et al., 1997), FbpC the NBP putative iron transporter from *Neisseria gonorrhoaea* (Lau et al., 2004) and MalK the NBP of the maltose transporter from *Salmonella typhimurium* (Schneider et al., 1995). In these studies it was found that including at least 2 mM of ATP and up to 50 % glycerol throughout the purification procedures greatly improved protein stability and prevented protein aggregation. This is possibly due to the requirement of NBP's to be stable with ATP bound and the effect of glycerol disrupting hydrophobic interactions. Therefore we decided to include 5 mM ATP and 20 % glycerol in the buffer during cell lysis and protein purification. However, in hindsight the high concentration of ATP and glycerol may have contributed towards contaminant co-elution and further work will require optimising the buffer.

Despite the low purity of *Mtb* UgpC we decided to investigate ATPase activity using the malachite green phosphate assay. Only a small number of studies have measured ATPase activity of prokaryote NBP's from ABC transporters (Nikaido et al., 1997) (Lau et al., 2004) (E. Schneider et al., 1995). However the studies found that addition of Mg^{2+} or alternative divalent metals including Co^{2+} and Mn^{2+} were required for optimal ATPase activity. Therefore we decided to include 1 mM $MgCl_2$ in the assay buffer. From the malachite green phosphate assay we found that phosphate release was higher in the presence *Mtb* UgpC and ATP compared to ATP without *Mtb* UgpC present, suggesting that *Mtb* UgpC did have some ATPase activity (6.3.18). As the

assay was only carried out once it would need to be repeated at least 3 times to confirm ATPase activity. However, as the *Mtb* UgpC protein preparation was impure and also contained a chaperone, *E. coli* DnaK that is known to hydrolyse ATP (Liberek et al., 1991) further purification is required to remove this contaminant. The assay also requires further optimisation because it was unclear why *Mtb* UgpC alone produced an increased absorbance signal despite extensive dialysis to remove the 5 mM ATP present in the purification buffer. Another problem may be due to high acidity of the assay liberating free phosphate from ATP in the absence of ATPase enzymatic activity. Instead it may be advantageous to use alternative methods to measure ATPase activity for example fluorescent methods have been reported (Brune et al., 1994) that may be more stable and sensitive compared to this colorimetric assay. Alternatively ADP release instead of phosphate release could be measured to report ATPase activity, using enzyme coupled systems and luminescent reagents (Zegzouti et al., 2009).

We also attempted to express and purify the predicted membrane spanning proteins *Mtb* UgpA and *Mtb* UgpE of the *Mtb* UgpABCE transporter. Being integral membrane proteins they are potentially more challenging to express because they may be toxic and more challenging to purify because detergents need to be added to stabilise and mimic the lipid environment (Schlegel et al., 2014). Expressing integral membrane proteins in the pWaldo vector in *E. coli* Lemo21 (DE3) cells and using a tuneable L-rhamnose gradient has previously found to be effective (Drew et al., 2001). Therefore we cloned the combined *Mtb ugpAE* genes into pWaldo and expressed it in this system, however, no expression was observed and only low GFP fluorescence was observed in all conditions tested. Instead a better system may have been to use the pDuet vector (Novagen) that has two multiple cloning sites containing different origins of replication to independently co-express *Mtb* UgpA and *Mtb* UgpE in the same plasmid. A previous study successfully expressed 70 *Mtb* integral membrane proteins in pET vectors (Korepanova et al., 2005). One of the strategies to express *Mtb* integral membrane proteins was to use the *E. coli* BL21 (DE3) Codon Plus-RP (Stratagene) strain that is designed to compensate for the codon usage differences of *Mtb* and *E. coli* (Korepanova et al., 2005). The difficulty of solving *Mtb* integral membrane protein structures is partly due to the difficulty in purifying sufficient quantities of protein. This is reflected by the observation of only a single X-ray structure of an integral membrane *Mtb* protein is available (PDB ID 2OAD), the *Mtb* mechanosensitive channels of large (MscL), reported at 3.5 Å resolution (G. Chang et al., 1998).

Previous studies have demonstrated it is possible to express and purify the membrane spanning proteins of ABC transporters (Wuttge et al., 2012). For example the *E. coli* $UgpAEC_2$ complex consisting of the two membrane spanning proteins and UgpC, the NBP were co-expressed in pET vectors under the control of the T7 promoter in *E coli*. The membrane protein complex was isolated from membranes and extracted with the detergent n-dodecyl-B-D-maltoside and purified by Ni^{2+} IMAC and gel filtration chromatography. The purified complex was re-constituted into proteoliposomes and was found to have ATPase activity (Wuttge et al., 2012). Therefore another more effective strategy may be to express the *Mtb* UgpA and *Mtb* UgpE the membrane spanning proteins together with the *Mtb* UgpC NBP as a complex. Also instead of using detergents for membrane protein purification, styrene maleic acid co-polymers (SMALPS) could be used to stabilise the complex because it would remain in the natural membrane environment (Gulati et al., 2014).

Previous studies of the *E. coli* maltose ABC transporter have been used to model the alternating access transport mechanism of type I ABC transporters (Locher, 2016). UgpABCE is also a predicted type I ABC transporter and therefore may undergo a similar transport mechanism to the maltose ABC transporter. This mechanism utilises ATP hydrolysis activity of the NBP dimer and substrate binding to the SBP in order to transport the substrate across the membrane. However, to prevent depleting the cell of ATP, the transport mechanism is closely coupled to substrate binding by the SBP so that ATP is not used in the absence of substrate.

Initially, the transporter exists in the inward facing conformation (Figure 6.29 (1)). The SBP present in the periplasm binds to its cognate substrate and adopts a closed conformation. The closed substrate bound SBP then engages with the external side of the transporter. Engagement of the SBP to the transporter triggers a conformational change that results in the NBP dimer closing (Figure 6.29 (2)). Because the NBP dimer is attached to the transmembrane domains, closure also converts the transporter to the outward facing conformation. A tunnel forms between the SBP and the transmembrane domains with the substrate bound in a temporary low affinity-binding site, half way across the membrane. In the final step, the closed NBP dimer hydrolyses ATP to ADP and inorganic phosphate resulting in a conformational change in the NBP to the inward facing conformation. This opens the transporter and enables the substrate to dissociate into the cytoplasm (Figure 6.29 (3)). In order for the cycle to reset, ADP is exchanged for ATP, priming the transporter to bind to the substrate bound SBP and the transport cycle is then repeated.

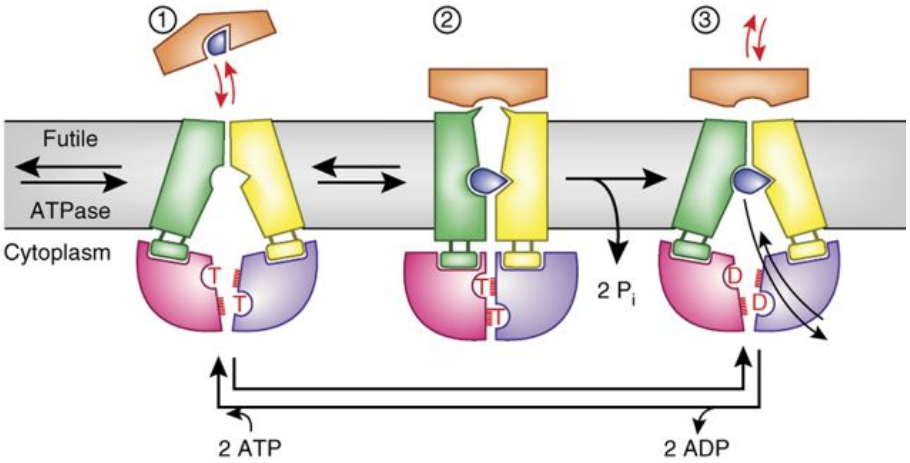


Figure 6.29: The predicted transport mechanism of the type I UgpABCE transporter (Locher, 2016). Green and yellow represent the transmembrane domains (TMD). Pink and purple represent the nucleotide binding proteins (NBP). Orange represents the substrate binding protein (SBP). Dark blue represents the substrates.

Therefore in the absence of the transmembrane domains, the UgpC NBP dimer in isolation is likely to be separated and not possess ATPase activity. A better experimental approach to investigate ATPase activity would be to express and purify the transmembrane domains together with the NBP and purify the membrane complex using the SMALP method.

Conclusions and future work

Conclusions

Overall the aim of this PhD was to carry out an extensive structural and functional analysis of the substrate-binding protein *Mtb* UgpB. The specific aims are described (Chapter 1, 1.3). Summaries of the key achievements from each chapter are described here.

Chapter 3: Cloning, expression and purification of *Mycobacterium tuberculosis* UgpB.

The substrate-binding protein, *Mtb ugpB* (Rv2833c) of the *Mtb ugpABCE* transporter was successfully expressed and purified to homogeneity. To do this, N-terminally truncated *Mtb* UgpB (residues 1-34) was expressed in pYUB1062 under the control of the acetamidase promoter in the more closely related host, *Mycobacterium smegmatis* $mc^2$4517. The expression yielded approximately 1.1 mg of protein per L culture and was sufficient to carry out further biochemical analyses. However, the purification strategy had to first be optimised due to the co-elution of several low molecular contaminants. By using a combination of IMAC, using Co^{2+} instead of Ni^{2+}, followed by QHP anion exchange chromatography and gel filtration chromatography using Superdex 75 instead of Superdex 200, we successfully purified *Mtb* UgpB. This was crucial to carry out a more extensive biochemical characterisation of ligand binding and for structural studies. Following successfully solving the X-ray structure of glycerol-3-phosphocholine (GPC) bound *Mtb* UgpB (Chapter 5, 5.3.7) we then generated site directed mutants based on the binding site and produced these genetic variants using the same strategy. The genetic variants were expressed and purified with yields between 0.2-2.6 mg of protein per L culture. All of the genetic variants were found to be soluble and stable and had similar secondary structure profiles as measured by circular dichroism analyses.

Chapter 4: Characterisation of *Mycobacterium tuberculosis* UgpB.

By bioinformatics, *Mtb* UgpB is predicted to be a carbohydrate binding protein therefore to carry out a more extensive biochemical characterisation we investigated binding of *Mtb* UgpB to a wide range of potential substrates. We measured binding to a panel carbohydrates, amino acids, glycerophosphodiester variations and anti-mycobacterial antibiotics by thermal shift assay and found GPC produced the highest thermal shift of +7.3 °C. This result was comparable to a previous study by isothermal calorimetry (Jiang et al., 2014). Our lipid strip analyses also found that *Mtb* UgpB binds to

phosphoinositides, phosphatidylserine and sulfatide, suggesting it may also have adapted to bind to alternative glycerophosphodiesers. Microscale thermophoresis (MST) was used to investigate the specificity of binding by testing GPC, the structurally related glycerol-3-phosphate (G3P) and phosphocholine. The MST results revealed that *Mtb* UgpB binds to GPC, with a K_d of 3.6 μM, suggesting that the combined glycerol, phosphate and choline moieties are required for binding recognition. By MST analysis all of the genetic variants of *Mtb* UgB had a lower binding affinity and some variants showed no binding for GPC compared to apo-*Mtb* UgpB. This confirmed that these residues, identified from the GPC bound *Mtb* UgpB X-ray structure are important for GPC binding recognition.

Based on the GPC bound structure (Chapter 5, 5.3.7) no interactions were found to occur through the choline head-group therefore we hypothesised that alternative glyerophosphodiesters containing different head-groups could also bind to *Mtb* UgpB. We attempted to produce glycerophosphodiesters using a synthetic route to produce glycerophosphoethanolamine (GPE) and glycerophosphoserine (GPS). We reacted protected building block reagents to synthesise these glycerophosphodiesters, however in all cases a mixture of the correct product, and by-products were produced, that could not be effectively separated. Therefore we decided to use a different route starting with the intact phospholipids containing the glycerophosphodiester, and removing the fatty acid tails using a chemical or enzymatic method. The chemical method resulted in several by-products however the enzymatic method produced the correct GPE and GPS products, and were characterised. We also produced a GPC enantiomer using the same enzymatic method to investigate if the stereo-configuration of the 2-hydroxyl of the glycerol group affected GPC binding recognition. Binding of the alternative glycerophosphodiesters was investigated by MST and revealed the order of binding preference to be: GPC>GPS>GPE>GPI(4)P>GPI>GPC enantiomer. Combined, this extensive biochemical characterisation suggested that *Mtb* UgpB is promiscuous for binding to a wide range of glycerophosphodiester substrates that *Mtb* may acquire during infection.

Chapter 5: Structural studies of *Mycobacterium tuberculosis* UgpB.

In order to co-crystallise *Mtb* UgpB and GPC we had to first reductively methylate the protein. We obtained crystals of *Mtb* UgpB and solved the structure of GPC bound *Mtb* UgpB by X-ray crystallography. The GPC bound structure revealed that *Mtb* UgpB had the typical structure of substrate-binding proteins (type II, cluster D) of the ATP-binding cassette (ABC) family. Upon GPC binding, *Mtb* UgpB had undergone closure between

the two α/β lobes. We also identified the precise molecular interactions in the GPC binding pocket, revealing that the glycerol and phosphate groups of GPC were the primary binding recognition elements. Residues Arg385 and Asp102 hydrogen bonded with the glycerol group. The glycerol backbone was also stabilised by π-stacking interactions from Trp208 and Van Der Waal's forces between Leu205. Tyr78, Tyr345, Ser153 and Ser272 hydrogen bonded with the phosphate moiety. Surprisingly from the structure no interactions were identified for the choline head-group despite the clear preference for binding to GPC. Our GPC bound X-ray structure observations were also supported by in-solution STD-NMR experiments in collaboration with Dr Ridvan Nepravishta and Dr Jesus Angulo, that revealed that the highest STD intensity signals involved the glycerol group. STD-NMR also revealed binding of the alternative glycerophisphodiester: glycerophosphoinositol-4-phosphate (GPI(4)P) and revealed that the glycerol backbone had re-orientated in order to accommodate the larger inositol-4-phosphate head-group. We also wanted to determine the binding mode of glycerophosphoserine (GPS) and we were able to solve a putative GPS bound *Mtb* UgpB structure by X-ray crystallography. The putative GPS bound structure revealed clear density for the glycerol and phosphate moieties, which aligned identically to the glycerol and phosphate moieties of GPC. However the serine head-group density was not as well defined.

Chapter 6: Cloning, expression and purification of *Mycobacterium tuberculosis* UgpC and UgpAE.

We also wanted to determine the structure and function of the other components of the *Mtb* UgpABCE transporter. We successfully expressed the nucleotide-binding protein *Mtb ugpC* (Rv2832c). To do this a range of different conditions were tested including: using different expression systems, host strains, IPTG concentrations, media additives and tags. We successfully expressed *Mtb* UgpC using the SUMO, solubility enhancement tag, which was confirmed by mass-spectrometry. ATPase activity was initially investigated using a malachite green phosphate assay however due to the low protein purity, ATPase activity could not be confirmed. We also investigated *ugpAE* the membrane spanning proteins and cloned these genes into the vector pWaldo. However, we could not successfully express these proteins using the *E. coli* Lemo21 (DE3) tuneable L-rhamnose expression system.

Future work

The work carried out during this PhD has raised several further questions that need to be addressed.

We have successfully revealed the molecular basis of *Mtb* UgpB binding to GPC and also found that *Mtb* UgpB is promiscuous for binding to alternative glycerophosphodiesters, potentially derived from macrophage membranes. Future work will therefore require further biochemical and structural characterisation of these substrates. We would especially like to establish how the binding recognition mode of GPS compares to GPC given that these substrates have very similar binding affinities. This will require further optimisation of the GPS co-crystallisation conditions. GPC was found to bind primarily through the glycerol and phosphate groups however it's still not clear why glycerol-3-phosphate cannot be recognised. Potentially differences in substrate head-group charge play a role therefore this could be investigated further by setting crystal trials and measuring binding at different pH values. Intriguingly, we also found ethambutol, the anti-TB drug, to have low binding to *Mtb* UgpB by thermal shift assay. Ethambutol is structurally similar to GPC as it also contains several hydroxyl groups therefore this potential binding requires further investigation.

Overall we want to determine the structure and function of the whole *Mtb* UgpABCE transporter, as this is the physiological unit, and may reveal important biochemical insights. We've had some success expressing the nucleotide-binding protein *Mtb* UgpC together with a SUMO tag. However we need to further optimise the expression and purification conditions in order to produce pure protein. This will enable us to investigate if *Mtb* UgpC is a functional ATPase. We also attempted to express the membrane spanning proteins, UgpA and UgpE. In the future, we could also try expressing the whole transporter complex together and extract the complex using SMALP's. Other strategies include independently co-expressing the membrane spanning proteins, *Mtb ugpA* and *Mtb ugpE* in the pDuet vector (Novagen). Different strains could also be used such as *E. coli* BL21 (DE3) Codon-PlusRP, optimised for efficient codon usage, or *E. coli* (DE3) ArcticExpress cells, designed to have improved protein expression at lower temperatures.

We now have an excellent understanding of the substrates of *Mtb* UgpB and how they are recognised therefore this provides an excellent basis to investigate the role of *Mtb* UgpB *in vivo*. In the lab, Magdalena Karlikowska has successfully produced an *Mtb* knockout of the whole UgpABCE transporter that is currently being evaluated in mice.

The *Mtb* UgpABCE knockout strain will also facilitate investigations of substrate uptake in whole cells and serve as a useful tool for comparisons to wild type. Alternatively, *Mycobacterium bovis* BCG could be used to investigate substrate uptake in whole cells. *M. bovis* BCG has a frame-shifted, non-functional *ugpB* gene (Brosch et al., 2007) therefore by cloning the functional *Mtb ugpB* gene into a vector such as pMV261 (Stover et al., 1991) for constitutive expression in *M. bovis* BCG, we could investigate substrate uptake and assess if the UgpABCE transporter function is restored. Several potential methods are available to investigate substrate uptake in whole cells. For example ion chromatography (Dionex) could be used to quantitatively measure substrate depletion from the culture media (Hanko et al., 2000) or radioactive or fluorescent reporter substrates could be generated and used to evaluate if glycerophosphodiesters are recycled and incorporated into the *Mtb* cell membrane.

Identification of new targets for drug development is essential to combat the emergence of drug resistant tuberculosis. Importantly, our structural analysis of glycerophosphodiester recognition by *Mtb* UgpB has facilitated the possibility of future medicinal chemistry research. Competitive inhibitors of *Mtb* UgpB could now be strategically designed to have tighter binding compared to GPC, leaving the transporter unable to function. Furthermore, given that it is really difficult to transport antibiotics across the *Mtb* cell wall, if we know what particular molecules are transported, then this could be used as a route to potentially smuggle toxic molecules into the cell. For example the GPC ligand could be modified to include a toxic head-group that when imported causes *Mtb* cell death. This could follow a similar “Trojan horse-like” strategy to previous ABC-transporter studies (Pletzer et al., 2014).

In conclusion, this PhD has opened up several new exciting avenues for research. Our biochemical and structural characterisation of *Mtb* UgpB has revealed the molecular basis of binding to GPC and alternative glycerophosphodiesters that requires further investigation. Recently, a potential *Mtb* phospholipid recycling pathway has been discovered (Larrouy-Maumus et al., 2013) therefore its tantalising to postulate that the *Mtb* UgpABCE transporter may play a crucial role in the uptake of glycerophosphodiesters, potentially derived from host macrophage membranes during infection.

Bibliography

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., . . . Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 2), 213-221. doi:10.1107/S0907444909052925

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., . . . Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr, 68*(Pt 4), 352-367. doi:10.1107/S0907444912001308

Aguado, J. M., Herrero, J. A., Gavalda, J., Torre-Cisneros, J., Blanes, M., Rufi, G., . . . Cantarell, C. (1997). Clinical presentation and outcome of tuberculosis in kidney, liver, and heart transplant recipients in Spain. Spanish Transplantation Infection Study Group, GESITRA. *Transplantation, 63*(9), 1278-1286. doi:10.1097/00007890-199705150-00015

Ai, J. W., Ruan, Q. L., Liu, Q. H., & Zhang, W. H. (2016). Updates on the risk factors for latent tuberculosis reactivation and their managements. *Emerg Microbes Infect, 5*, e10. doi:10.1038/emi.2016.10

Alderwick, L. J. (2019). https://www.birmingham.ac.uk/staff/profiles/biosciences/alderwick-luke.aspx.

Alderwick, L. J., Birch, H. L., Mishra, A. K., Eggeling, L., & Besra, G. S. (2007). Structure, function and biosynthesis of the Mycobacterium tuberculosis cell wall: arabinogalactan and lipoarabinomannan assembly with a view to discovering new drug targets. *Biochem Soc Trans, 35*(Pt 5), 1325-1328. doi:10.1042/BST0351325

Alderwick, L. J., Harrison, J., Lloyd, G. S., & Birch, H. L. (2015). The Mycobacterial Cell Wall--Peptidoglycan and Arabinogalactan. *Cold Spring Harb Perspect Med, 5*(8), a021113. doi:10.1101/cshperspect.a021113

Alderwick, L. J., Seidel, M., Sahm, H., Besra, G. S., & Eggeling, L. (2006). Identification of a novel arabinofuranosyltransferase (AftA) involved in cell wall arabinan biosynthesis in Mycobacterium tuberculosis. *J Biol Chem, 281*(23), 15653-15661. doi:10.1074/jbc.M600045200

Andries, K., Verhasselt, P., Guillemont, J., Gohlmann, H. W., Neefs, J. M., Winkler, H., . . . Jarlier, V. (2005). A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. *Science, 307*(5707), 223-227. doi:10.1126/science.1106753

Apt, A., & Kramnik, I. (2009). Man and mouse TB: contradictions and solutions. *Tuberculosis (Edinb), 89*(3), 195-198. doi:10.1016/j.tube.2009.02.002

Atkins, D., al-Ghusein, H., Prehaud, C., & Coates, A. R. (1994). Overproduction and purification of Mycobacterium tuberculosis chaperonin 10. *Gene, 150*(1), 145-148.

Aubry, A., Chosidow, O., Caumes, E., Robert, J., & Cambau, E. (2002). Sixty-three cases of Mycobacterium marinum infection: clinical features, treatment, and antibiotic susceptibility of causative isolates. *Arch Intern Med, 162*(15), 1746-1752. doi:10.1001/archinte.162.15.1746

Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K. S., Wilson, T., . . . Jacobs, W. R., Jr. (1994). inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. *Science, 263*(5144), 227-230. doi:10.1126/science.8284673

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. *Curr Opin Biotechnol, 10*(5), 411-421.

Bang, H. J., Kim, I. H., & Kim, B. H. (2016). Phospholipase A1-catalyzed hydrolysis of soy phosphatidylcholine to prepare l-alpha-glycerylphosphorylcholine in organic-aqueous media. *Food Chem, 190*, 201-206. doi:10.1016/j.foodchem.2015.05.093

Bansal-Mutalik, R., & Nikaido, H. (2014). Mycobacterial outer membrane is a lipid bilayer and the inner membrane is unusually rich in diacyl phosphatidylinositol dimannosides. *Proc Natl Acad Sci U S A, 111*(13), 4958-4963. doi:10.1073/pnas.1403078111

Barberis, I., Bragazzi, N. L., Galluzzo, L., & Martini, M. (2017). The history of tuberculosis: from the first historical records to the isolation of Koch's bacillus. *J Prev Med Hyg, 58*(1), E9-E12.

Bashiri, G., & Baker, E. N. (2015). Production of recombinant proteins in Mycobacterium smegmatis for structural and functional studies. *Protein Sci, 24*(1), 1-10. doi:10.1002/pro.2584

Bashiri, G., Perkowski, E. F., Turner, A. P., Feltcher, M. E., Braunstein, M., & Baker, E. N. (2012). Tat-dependent translocation of an F420-binding protein of Mycobacterium tuberculosis. *PLoS One, 7*(10), e45003. doi:10.1371/journal.pone.0045003

Baynes, B. M., Wang, D. I., & Trout, B. L. (2005). Role of arginine in the stabilization of proteins against aggregation. *Biochemistry, 44*(12), 4919-4925. doi:10.1021/bi047528r

Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S., & Small, P. M. (1999). Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science, 284*(5419), 1520-1523. doi:10.1126/science.284.5419.1520

Belanger, A. E., Besra, G. S., Ford, M. E., Mikusova, K., Belisle, J. T., Brennan, P. J., & Inamine, J. M. (1996). The embAB genes of Mycobacterium avium encode an arabinosyl transferase involved in cell wall arabinan biosynthesis that is the target for the antimycobacterial drug ethambutol. *Proc Natl Acad Sci U S A, 93*(21), 11919-11924. doi:10.1073/pnas.93.21.11919

Belisle, J. T., Vissa, V. D., Sievert, T., Takayama, K., Brennan, P. J., & Besra, G. S. (1997). Role of the major antigen of Mycobacterium tuberculosis in cell wall biogenesis. *Science, 276*(5317), 1420-1422. doi:10.1126/science.276.5317.1420

Bellinzoni, M., De Rossi, E., Branzoni, M., Milano, A., Peverali, F. A., Rizzi, M., & Riccardi, G. (2002). Heterologous expression, purification, and enzymatic activity of Mycobacterium tuberculosis NAD(+) synthetase. *Protein Expr Purif, 25*(3), 547-557.

Bellinzoni, M., & Riccardi, G. (2003). Techniques and applications: The heterologous expression of Mycobacterium tuberculosis genes is an uphill road. *Trends Microbiol, 11*(8), 351-358.

Benard, E. L., van der Sar, A. M., Ellett, F., Lieschke, G. J., Spaink, H. P., & Meijer, A. H. (2012). Infection of zebrafish embryos with intracellular bacterial pathogens. *J Vis Exp*(61). doi:10.3791/3781

Benjak, A., Uplekar, S., Zhang, M., Piton, J., Cole, S. T., & Sala, C. (2016). Genomic and transcriptomic analysis of the streptomycin-dependent Mycobacterium tuberculosis strain 18b. *BMC Genomics, 17*, 190. doi:10.1186/s12864-016-2528-2

Berntsson, R. P., Smits, S. H., Schmitt, L., Slotboom, D. J., & Poolman, B. (2010). A structural classification of substrate-binding proteins. *FEBS Lett, 584*(12), 2606-2617. doi:10.1016/j.febslet.2010.04.043

Besra, G. S., & Brennan, P. J. (1997). The mycobacterial cell wall: biosynthesis of arabinogalactan and lipoarabinomannan. *Biochem Soc Trans, 25*(3), 845-850. doi:10.1042/bst0250845

Besra, G. S., Morehouse, C. B., Rittner, C. M., Waechter, C. J., & Brennan, P. J. (1997). Biosynthesis of mycobacterial lipoarabinomannan. *J Biol Chem, 272*(29), 18460-18466. doi:10.1074/jbc.272.29.18460

Bhattacharyya, N., Nkumama, I. N., Newland-Smith, Z., Lin, L. Y., Yin, W., Cullen, R. E., . . . O'Hare, H. M. (2018). An Aspartate-Specific Solute-Binding Protein Regulates Protein Kinase G Activity To Control Glutamate Metabolism in Mycobacteria. *MBio, 9*(4). doi:10.1128/mBio.00931-18

Bigi, F., Taboga, O., Romano, M. I., Alito, A., Fisanotti, J. C., & Cataldi, A. A. (1999). Expression of the Mycobacterium bovis P36 gene in Mycobacterium smegmatis and the baculovirus/insect cell system. *Braz J Med Biol Res, 32*(1), 29-37. doi:10.1590/s0100-879x1999000100004

Billeter, M., Wagner, G., & Wuthrich, K. (2008). Solution NMR structure determination of proteins revisited. *J Biomol NMR, 42*(3), 155-158. doi:10.1007/s10858-008-9277-8

Bird, L. E., Rada, H., Verma, A., Gasper, R., Birch, J., Jennions, M., . . . Owens, R. J. (2015). Green fluorescent protein-based expression screening of membrane proteins in Escherichia coli. *J Vis Exp*(95), e52357. doi:10.3791/52357

Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., . . . Shao, Y. (1997). The complete genome sequence of Escherichia coli K-12. *Science, 277*(5331), 1453-1462.

Blaum, B. S., Neu, U., Peters, T., & Stehle, T. (2018). Spin ballet for sweet encounters: saturation-transfer difference NMR and X-ray crystallography complement each other in the elucidation of protein-glycan interactions. *Acta Crystallogr F Struct Biol Commun, 74*(Pt 8), 451-462. doi:10.1107/S2053230X18006581

BMJ. (1955). Various combinations of isoniazid with streptomycin or P.A.S in the treatment of pulmonary tuberculosis. *British Medical Journal*.

Bornhorst, J. A., & Falke, J. J. (2000). Purification of proteins using polyhistidine affinity tags. *Methods Enzymol, 326*, 245-254. doi:10.1016/s0076-6879(00)26058-8

Braibant, M., Gilot, P., & Content, J. (2000). The ATP binding cassette (ABC) transport systems of Mycobacterium tuberculosis. *FEMS Microbiol Rev, 24*(4), 449-467. doi:10.1111/j.1574-6976.2000.tb00550.x

Brautigam, C. A., Ouyang, Z., Deka, R. K., & Norgard, M. V. (2014). Sequence, biophysical, and structural analyses of the PstS lipoprotein (BB0215) from Borrelia burgdorferi reveal a likely binding component of an ABC-type phosphate transporter. *Protein Sci, 23*(2), 200-212. doi:10.1002/pro.2406

Brennan, P. J., & Nikaido, H. (1995). The envelope of mycobacteria. *Annu Rev Biochem, 64*, 29-63. doi:10.1146/annurev.bi.64.070195.000333

Briken, V., & Miller, J. L. (2008). Living on the edge: inhibition of host cell apoptosis by Mycobacterium tuberculosis. *Future Microbiol, 3*(4), 415-422. doi:10.2217/17460913.3.4.415

Briken, V., Porcelli, S. A., Besra, G. S., & Kremer, L. (2004). Mycobacterial lipoarabinomannan and related lipoglycans: from biogenesis to modulation

of the immune response. *Mol Microbiol, 53*(2), 391-403. doi:10.1111/j.1365-2958.2004.04183.x

Brites, D., & Gagneux, S. (2012). Old and new selective pressures on Mycobacterium tuberculosis. *Infect Genet Evol, 12*(4), 678-685. doi:10.1016/j.meegid.2011.08.010

Brites, D., & Gagneux, S. (2015). Co-evolution of Mycobacterium tuberculosis and Homo sapiens. *Immunol Rev, 264*(1), 6-24. doi:10.1111/imr.12264

Brosch, R., Gordon, S. V., Garnier, T., Eiglmeier, K., Frigui, W., Valenti, P., . . . Cole, S. T. (2007). Genome plasticity of BCG and impact on vaccine efficacy. *Proc Natl Acad Sci U S A, 104*(13), 5596-5601. doi:10.1073/pnas.0700869104

Brosch, R., Gordon, S. V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., . . . Cole, S. T. (2002). A new evolutionary scenario for the Mycobacterium tuberculosis complex. *Proc Natl Acad Sci U S A, 99*(6), 3684-3689. doi:10.1073/pnas.052548299

Brune, M., Hunter, J. L., Corrie, J. E., & Webb, M. R. (1994). Direct, real-time measurement of rapid inorganic phosphate release using a novel fluorescent probe and its application to actomyosin subfragment 1 ATPase. *Biochemistry, 33*(27), 8262-8271. doi:10.1021/bi00193a013

Brust, B., Lecoufle, M., Tuaillon, E., Dedieu, L., Canaan, S., Valverde, V., & Kremer, L. (2011). Mycobacterium tuberculosis lipolytic enzymes as potential biomarkers for the diagnosis of active tuberculosis. *PLoS One, 6*(9), e25078. doi:10.1371/journal.pone.0025078

Burke, J. E., & Dennis, E. A. (2009). Phospholipase A2 structure/function, mechanism, and signaling. *J Lipid Res, 50 Suppl*, S237-242. doi:10.1194/jlr.R800033-JLR200

Cadena, A. M., Fortune, S. M., & Flynn, J. L. (2017). Heterogeneity in tuberculosis. *Nat Rev Immunol, 17*(11), 691-702. doi:10.1038/nri.2017.69

Callahan, M. K., Williamson, P., & Schlegel, R. A. (2000). Surface expression of phosphatidylserine on macrophages is required for phagocytosis of apoptotic thymocytes. *Cell Death Differ, 7*(7), 645-653. doi:10.1038/sj.cdd.4400690

Calmette, A. (1922). The Protection of Mankind against Tuberculosis: Being an Address before the Medico-Chirurgical Society of Edinburgh. *Edinb Med J, 29*(1), 93-104.

Cambier, C. J., Falkow, S., & Ramakrishnan, L. (2014). Host evasion and exploitation schemes of Mycobacterium tuberculosis. *Cell, 159*(7), 1497-1509. doi:10.1016/j.cell.2014.11.024

Cambier, C. J., Takaki, K. K., Larson, R. P., Hernandez, R. E., Tobin, D. M., Urdahl, K. B., . . . Ramakrishnan, L. (2014). Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids. *Nature, 505*(7482), 218-222. doi:10.1038/nature12799

Chandravanshi, M., Gogoi, P., & Kanaujia, S. P. (2016). Computational characterization of TTHA0379: A potential glycerophosphocholine binding protein of Ugp ATP-binding cassette transporter. *Gene, 592*(2), 260-268. doi:10.1016/j.gene.2016.07.017

Chang, G., Spencer, R. H., Lee, A. T., Barclay, M. T., & Rees, D. C. (1998). Structure of the MscL homolog from Mycobacterium tuberculosis: a gated mechanosensitive ion channel. *Science, 282*(5397), 2220-2226. doi:10.1126/science.282.5397.2220

Chang, J. C., Miner, M. D., Pandey, A. K., Gill, W. P., Harik, N. S., Sassetti, C. M., & Sherman, D. R. (2009). igr Genes and Mycobacterium tuberculosis

cholesterol metabolism. *J Bacteriol, 191*(16), 5232-5239. doi:10.1128/JB.00452-09

Chao, M. C., & Rubin, E. J. (2010). Letting sleeping dos lie: does dormancy play a role in tuberculosis? *Annu Rev Microbiol, 64*, 293-311. doi:10.1146/annurev.micro.112408.134043

Chen, J., Zhang, S., Cui, P., Shi, W., Zhang, W., & Zhang, Y. (2017). Identification of novel mutations associated with cycloserine resistance in Mycobacterium tuberculosis. *J Antimicrob Chemother, 72*(12), 3272-3276. doi:10.1093/jac/dkx316

Cimmperman, P., Baranauskiene, L., Jachimoviciute, S., Jachno, J., Torresan, J., Michailoviene, V., . . . Matulis, D. (2008). A quantitative model of thermal stabilization and destabilization of proteins by ligands. *Biophys J, 95*(7), 3222-3231. doi:10.1529/biophysj.108.134973

Clarke, T. E., Braun, V., Winkelmann, G., Tari, L. W., & Vogel, H. J. (2002). X-ray crystallographic structures of the Escherichia coli periplasmic protein FhuD bound to hydroxamate-type siderophores and the antibiotic albomycin. *J Biol Chem, 277*(16), 13966-13972. doi:10.1074/jbc.M109385200

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., . . . Barrell, B. G. (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature, 393*(6685), 537-544. doi:10.1038/31159

Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr, 50*(Pt 5), 760-763. doi:10.1107/S0907444994003112

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., . . . Gagneux, S. (2013). Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nat Genet, 45*(10), 1176-1182. doi:10.1038/ng.2744

Comstock, G. W. (2008). Frost revisited: the modern epidemiology of tuberculosis: the third Wade Hampton Frost Lecture. *Am J Epidemiol, 168*(7), 692-711. doi:10.1093/aje/kwn268

Concepcion, J., Witte, K., Wartchow, C., Choo, S., Yao, D., Persson, H., . . . Tan, H. (2009). Label-free detection of biomolecular interactions using BioLayer interferometry for kinetic characterization. *Comb Chem High Throughput Screen, 12*(8), 791-800.

Cooper, A. M. (2014). Mouse model of tuberculosis. *Cold Spring Harb Perspect Med, 5*(2), a018556. doi:10.1101/cshperspect.a018556

Cooper, J. B., McIntyre, K., Badasso, M. O., Wood, S. P., Zhang, Y., Garbe, T. R., & Young, D. (1995). X-ray structure analysis of the iron-dependent superoxide dismutase from Mycobacterium tuberculosis at 2.0 Angstroms resolution reveals novel dimer-dimer interactions. *J Mol Biol, 246*(4), 531-544. doi:10.1006/jmbi.1994.0105

Crubezy, E., Ludes, B., Poveda, J. D., Clayton, J., Crouau-Roy, B., & Montagnon, D. (1998). Identification of Mycobacterium DNA in an Egyptian Pott's disease of 5,400 years old. *C R Acad Sci III, 321*(11), 941-951.

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y., & McPherson, A. (1994). Screening and optimization strategies for macromolecular crystal growth. *Acta Crystallogr D Biol Crystallogr, 50*(Pt 4), 414-423. doi:10.1107/S0907444994002660

Cuneo, M. J., Changela, A., Warren, J. J., Beese, L. S., & Hellinga, H. W. (2006). The crystal structure of a thermophilic glucose binding protein reveals

adaptations that interconvert mono and di-saccharide binding sites. *J Mol Biol, 362*(2), 259-270. doi:10.1016/j.jmb.2006.06.084

Cuthbertson, J. M., Doyle, D. A., & Sansom, M. S. (2005). Transmembrane helix prediction: a comparative evaluation and analysis. *Protein Eng Des Sel, 18*(6), 295-308. doi:10.1093/protein/gzi032

Daffe, M., & Draper, P. (1998). The envelope layers of mycobacteria with reference to their pathogenicity. *Adv Microb Physiol, 39*, 131-203.

Daffe, M., & Etienne, G. (1999). The capsule of Mycobacterium tuberculosis and its implications for pathogenicity. *Tuber Lung Dis, 79*(3), 153-169. doi:10.1054/tuld.1998.0200

Daugelat, S., Kowall, J., Mattow, J., Bumann, D., Winter, R., Hurwitz, R., & Kaufmann, S. H. (2003). The RD1 proteins of Mycobacterium tuberculosis: expression in Mycobacterium smegmatis and biochemical characterization. *Microbes Infect, 5*(12), 1082-1095.

Davis, I. W., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res, 32*(Web Server issue), W615-619. doi:10.1093/nar/gkh398

De Matteis, M. A., & Godi, A. (2004). PI-loting membrane traffic. *Nat Cell Biol, 6*(6), 487-492. doi:10.1038/ncb0604-487

DeBarber, A. E., Mdluli, K., Bosman, M., Bekker, L. G., & Barry, C. E., 3rd. (2000). Ethionamide activation and sensitivity in multidrug-resistant Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A, 97*(17), 9677-9682. doi:10.1073/pnas.97.17.9677

Defacque, H., Bos, E., Garvalov, B., Barret, C., Roy, C., Mangeat, P., . . . Griffiths, G. (2002). Phosphoinositides regulate membrane-dependent actin assembly by latex bead phagosomes. *Mol Biol Cell, 13*(4), 1190-1202. doi:10.1091/mbc.01-06-0314

DEFRA. (2014). The Strategy for achieving Officially Bovine Tuberculosis Free status for England.

DeJesus, M. A., Gerrick, E. R., Xu, W., Park, S. W., Long, J. E., Boutte, C. C., . . . Ioerger, T. R. (2017). Comprehensive Essentiality Analysis of the Mycobacterium tuberculosis Genome via Saturating Transposon Mutagenesis. *MBio, 8*(1). doi:10.1128/mBio.02133-16

Dennis, E. A. (2015). Introduction to Thematic Review Series: Phospholipases: Central Role in Lipid Signaling and Disease. *J Lipid Res, 56*(7), 1245-1247. doi:10.1194/jlr.E061101

Deretic, V., Singh, S., Master, S., Harris, J., Roberts, E., Kyei, G., . . . Vergne, I. (2006). Mycobacterium tuberculosis inhibition of phagolysosome biogenesis and autophagy as a host defence mechanism. *Cell Microbiol, 8*(5), 719-727. doi:10.1111/j.1462-5822.2006.00705.x

Deuschle, K., Okumoto, S., Fehr, M., Looger, L. L., Kozhukh, L., & Frommer, W. B. (2005). Construction and optimization of a family of genetically encoded metabolite sensors by semirational protein engineering. *Protein Sci, 14*(9), 2304-2314. doi:10.1110/ps.051508105

Donnelly, C. A., Wei, G., Johnston, W. T., Cox, D. R., Woodroffe, R., Bourne, F. J., . . . Morrison, W. I. (2007). Impacts of widespread badger culling on cattle tuberculosis: concluding analyses from a large-scale field trial. *Int J Infect Dis, 11*(4), 300-308. doi:10.1016/j.ijid.2007.04.001

Donnelly, C. A., Woodroffe, R., Cox, D. R., Bourne, F. J., Cheeseman, C. L., Clifton-Hadley, R. S., . . . Morrison, W. I. (2006). Positive and negative effects of

widespread badger culling on tuberculosis in cattle. *Nature, 439*(7078), 843-846. doi:10.1038/nature04454

Donoghue, H. D., Lee, O. Y., Minnikin, D. E., Besra, G. S., Taylor, J. H., & Spigelman, M. (2010). Tuberculosis in Dr Granville's mummy: a molecular re-examination of the earliest known Egyptian mummy to be scientifically examined and given a medical diagnosis. *Proc Biol Sci, 277*(1678), 51-56. doi:10.1098/rspb.2009.1484

Downes, C. P., Gray, A., & Lucocq, J. M. (2005). Probing phosphoinositide functions in signaling and membrane trafficking. *Trends Cell Biol, 15*(5), 259-268. doi:10.1016/j.tcb.2005.03.008

Drew, D. E., von Heijne, G., Nordlund, P., & de Gier, J. W. (2001). Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. *FEBS Lett, 507*(2), 220-224.

Dubnau, E., Chan, J., Mohan, V. P., & Smith, I. (2005). responses of mycobacterium tuberculosis to growth in the mouse lung. *Infect Immun, 73*(6), 3754-3757. doi:10.1128/IAI.73.6.3754-3757.2005

Eastwood, J. B., Corbishley, C. M., & Grange, J. M. (2001). Tuberculosis and the kidney. *J Am Soc Nephrol, 12*(6), 1307-1314.

Edmondson, S. P. (1992). NOE RFactors and Structural Refinement Using FIRM, an Iterative Relaxation Matrix Program. *Journal of Magnetic Resonance, 98*, 283-298.

Edson, N. L. (1951). The intermediary metabolism of the mycobacteria. *Bacteriol Rev, 15*(3), 147-182.

Ehlers, S., & Schaible, U. E. (2012). The granuloma in tuberculosis: dynamics of a host-pathogen collusion. *Front Immunol, 3*, 411. doi:10.3389/fimmu.2012.00411

Ehrt, S., & Schnappinger, D. (2009). Mycobacterial survival strategies in the phagosome: defence against host stresses. *Cell Microbiol, 11*(8), 1170-1178. doi:10.1111/j.1462-5822.2009.01335.x

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr, 60*(Pt 12 Pt 1), 2126-2132. doi:10.1107/S0907444904019158

Eoh, H., & Rhee, K. Y. (2013). Multifunctional essentiality of succinate metabolism in adaptation to hypoxia in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A, 110*(16), 6554-6559. doi:10.1073/pnas.1219375110

Etienne, G., Laval, F., Villeneuve, C., Dinadayala, P., Abouwarda, A., Zerbib, D., . . . Daffe, M. (2005). The cell envelope structure and properties of Mycobacterium smegmatis mc(2)155: is there a clue for the unique transformability of the strain? *Microbiology, 151*(Pt 6), 2075-2086. doi:10.1099/mic.0.27869-0

FDA. (2019). FDA Approves New Treatment for Highly Drug-Resistant Forms of Tuberculosis. *TB Alliance*.

Fedrizzi, T., Meehan, C. J., Grottola, A., Giacobazzi, E., Fregni Serpini, G., Tagliazucchi, S., . . . Segata, N. (2017). Genomic characterization of Nontuberculous Mycobacteria. *Sci Rep, 7*, 45258. doi:10.1038/srep45258

Fennelly, K. P., Martyny, J. W., Fulton, K. E., Orme, I. M., Cave, D. M., & Heifets, L. B. (2004). Cough-generated aerosols of Mycobacterium tuberculosis: a new method to study infectiousness. *Am J Respir Crit Care Med, 169*(5), 604-609. doi:10.1164/rccm.200308-1101OC

Fessler, M. B., & Summer, R. S. (2016). Surfactant Lipids at the Host-Environment Interface. Metabolic Sensors, Suppressors, and Effectors of Inflammatory

Lung Disease. *Am J Respir Cell Mol Biol, 54*(5), 624-635. doi:10.1165/rcmb.2016-0011PS

Fox, W., Ellard, G. A., & Mitchison, D. A. (1999). Studies on the treatment of tuberculosis undertaken by the British Medical Research Council tuberculosis units, 1946-1986, with relevant subsequent publications. *Int J Tuberc Lung Dis, 3*(10 Suppl 2), S231-279.

Fox, W., & Sutherland, I. (1955). The clinical significance of positive cultures and of isoniazid-resistant tubercle bacilli during the treatment of pulmonary tuberculosis; report to the Tuberculosis Chemotherapy Trials Committee of the Medical Research Council. *Thorax, 10*(2), 85-98. doi:10.1136/thx.10.2.85

Fox, W., & Sutherland, I. (1956). A five-year assessment of patients in a controlled trial of streptomycin, para-aminosalicylic acid, and streptomycin plus para-aminosalicylic acid, in pulmonary tuberculosis. *Q J Med, 25*(98), 221-243.

Fox, W., Sutherland, I., & Daniels, M. (1954). A five-year assessment of patients in a controlled trial of streptomycin in pulmonary tuberculosis; report to the Tuberculosis Chemotherapy Trials Committee of the Medical Research Council. *Q J Med, 23*(91), 347-366.

Frieden, T. R., Sterling, T., Pablos-Mendez, A., Kilburn, J. O., Cauthen, G. M., & Dooley, S. W. (1993). The emergence of drug-resistant tuberculosis in New York City. *N Engl J Med, 328*(8), 521-526. doi:10.1056/NEJM199302253280801

Fukami-Kobayashi, K., Tateno, Y., & Nishikawa, K. (1999). Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *J Mol Biol, 286*(1), 279-290. doi:10.1006/jmbi.1998.2454

Fullam, E., Prokes, I., Futterer, K., & Besra, G. S. (2016). Structural and functional analysis of the solute-binding protein UspC from Mycobacterium tuberculosis that is specific for amino sugars. *Open Biol, 6*(6). doi:10.1098/rsob.160105

Gagneux, S. (2018). Ecology and evolution of Mycobacterium tuberculosis. *Nat Rev Microbiol, 16*(4), 202-213. doi:10.1038/nrmicro.2018.8

Gagnon, M., Sébastien, D., Bertrand, X., Auger, M., & Paquin, J. (2017). A Flexible Synthetic Approach to Phosphatidylglycerols. *European Journal of Organic Chemistry, 43*, 6401-6407.

Garg, R. K. (1999). Tuberculosis of the central nervous system. *Postgrad Med J, 75*(881), 133-140. doi:10.1136/pgmj.75.881.133

Garg, R. K., & Somvanshi, D. S. (2011). Spinal tuberculosis: a review. *J Spinal Cord Med, 34*(5), 440-454. doi:10.1179/2045772311Y.0000000023

Garnier, T., Eiglmeier, K., Camus, J. C., Medina, N., Mansoor, H., Pryor, M., . . . Hewinson, R. G. (2003). The complete genome sequence of Mycobacterium bovis. *Proc Natl Acad Sci U S A, 100*(13), 7877-7882. doi:10.1073/pnas.1130426100

Gawad, J., & Bonde, C. (2018). Decaprenyl-phosphoryl-ribose 2'-epimerase (DprE1): challenging target for antitubercular drug discovery. *Chem Cent J, 12*(1), 72. doi:10.1186/s13065-018-0441-2

Getahun, H., Chaisson, R. E., & Raviglione, M. (2015). Latent Mycobacterium tuberculosis Infection. *N Engl J Med, 373*(12), 1179-1180. doi:10.1056/NEJMc1508223

Goins, C. M., Schreidah, C. M., Dajnowicz, S., & Ronning, D. R. (2018). Structural basis for lipid binding and mechanism of the Mycobacterium tuberculosis

Rv3802 phospholipase. *J Biol Chem, 293*(4), 1363-1372. doi:10.1074/jbc.RA117.000240

Goldstein, B. P. (2014). Resistance to rifampicin: a review. *J Antibiot (Tokyo), 67*(9), 625-630. doi:10.1038/ja.2014.107

Goldstone, R. M., Moreland, N. J., Bashiri, G., Baker, E. N., & Shaun Lott, J. (2008). A new Gateway vector and expression protocol for fast and efficient recombinant protein expression in Mycobacterium smegmatis. *Protein Expr Purif, 57*(1), 81-87. doi:10.1016/j.pep.2007.08.015

Goren, M. B., Brokl, O., & Das, B. C. (1976). Sulfatides of Mycobacterium tuberculosis: the structure of the principal sulfatide (SL-I). *Biochemistry, 15*(13), 2728-2735. doi:10.1021/bi00658a003

Goren, M. B., Brokl, O., & Schaefer, W. B. (1974). Lipids of putative relevance to virulence in Mycobacterium tuberculosis: phthiocerol dimycocerosate and the attenuation indicator lipid. *Infect Immun, 9*(1), 150-158.

Goyal, S., Klassert, T. E., & Slevogt, H. (2016). C-type lectin receptors in tuberculosis: what we know. *Med Microbiol Immunol, 205*(6), 513-535. doi:10.1007/s00430-016-0470-1

Grange, J. M. (2001). Mycobacterium bovis infection in human beings. *Tuberculosis (Edinb), 81*(1-2), 71-77. doi:10.1054/tube.2000.0263

Griffith, D. E., & Kerr, C. M. (1996). Tuberculosis: disease of the past, disease of the present. *J Perianesth Nurs, 11*(4), 240-245.

Groschel, M. I., Sayes, F., Simeone, R., Majlessi, L., & Brosch, R. (2016). ESX secretion systems: mycobacterial evolution to counter host immunity. *Nat Rev Microbiol, 14*(11), 677-691. doi:10.1038/nrmicro.2016.131

Grosshennig, S., Schmidl, S. R., Schmeisky, G., Busse, J., & Stulke, J. (2013). Implication of glycerol and phospholipid transporters in Mycoplasma pneumoniae growth and virulence. *Infect Immun, 81*(3), 896-904. doi:10.1128/IAI.01212-12

Grumbach, F. (1970). [Activity of rifampicin on experimental tuberculosis in mice. The development of resistance to rifampicin. Therapeutic effects of combinations of different drugs with rifampicin]. *Antibiot Chemother, 16*, 392-405.

Gudmand-Hoyer, E., & Skovbjerg, H. (1996). Disaccharide digestion and maldigestion. *Scand J Gastroenterol Suppl, 216*, 111-121.

Gulati, S., Jamshad, M., Knowles, T. J., Morrison, K. A., Downing, R., Cant, N., . . . Rothnie, A. J. (2014). Detergent-free purification of ABC (ATP-binding-cassette) transporters. *Biochem J, 461*(2), 269-278. doi:10.1042/BJ20131477

Gupta, P., Aggarwal, N., Batra, P., Mishra, S., & Chaudhuri, T. K. (2006). Co-expression of chaperonin GroEL/GroES enhances in vivo folding of yeast mitochondrial aconitase and alters the growth characteristics of Escherichia coli. *Int J Biochem Cell Biol, 38*(11), 1975-1985. doi:10.1016/j.biocel.2006.05.013

Gupta, U. D., & Katoch, V. M. (2009). Animal models of tuberculosis for vaccine development. *Indian J Med Res, 129*(1), 11-18.

Gurvitz, A., Hiltunen, J. K., & Kastaniotis, A. J. (2009). Heterologous expression of mycobacterial proteins in Saccharomyces cerevisiae reveals two physiologically functional 3-hydroxyacyl-thioester dehydratases, HtdX and HtdY, in addition to HadABC and HtdZ. *J Bacteriol, 191*(8), 2683-2690. doi:10.1128/JB.01046-08

Gutierrez, M. C., Brisse, S., Brosch, R., Fabre, M., Omais, B., Marmiesse, M., . . . Vincent, V. (2005). Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. *PLoS Pathog, 1*(1), e5. doi:10.1371/journal.ppat.0010005

Hancox, M. (2002). Bovine tuberculosis: milk and meat safety. *Lancet, 359*(9307), 706-707. doi:10.1016/S0140-6736(02)07786-3

Hanko, V. P., & Rohrer, J. S. (2000). Determination of carbohydrates, sugar alcohols, and glycols in cell cultures and fermentation broths using high-performance anion-exchange chromatography with pulsed amperometric detection. *Anal Biochem, 283*(2), 192-199. doi:10.1006/abio.2000.4653

Harries, A. D., Lin, Y., Satyanarayana, S., Lonnroth, K., Li, L., Wilson, N., . . . Kapur, A. (2011). The looming epidemic of diabetes-associated tuberculosis: learning lessons from HIV-associated tuberculosis. *Int J Tuberc Lung Dis, 15*(11), 1436-1444, i. doi:10.5588/ijtld.11.0503

Havlir, D. V., & Barnes, P. F. (1999). Tuberculosis in patients with human immunodeficiency virus infection. *N Engl J Med, 340*(5), 367-373. doi:10.1056/NEJM199902043400507

Hayward, S., & Lee, R. A. (2002). Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J Mol Graph Model, 21*(3), 181-183.

Heras, B., Edeling, M. A., Byriel, K. A., Jones, A., Raina, S., & Martin, J. L. (2003). Dehydration converts DsbG crystal diffraction from low to high resolution. *Structure, 11*(2), 139-145.

Hill, P. C., Jackson-Sillah, D., Donkor, S. A., Otu, J., Adegbola, R. A., & Lienhardt, C. (2006). Risk factors for pulmonary tuberculosis: a clinic-based case control study in The Gambia. *BMC Public Health, 6*, 156. doi:10.1186/1471-2458-6-156

Hirsch, C. S., Ellner, J. J., Russell, D. G., & Rich, E. A. (1994). Complement receptor-mediated uptake and tumor necrosis factor-alpha-mediated growth inhibition of Mycobacterium tuberculosis by human alveolar macrophages. *J Immunol, 152*(2), 743-753.

Hobby, G. L., & Lenert, T. F. (1957). The in vitro action of antituberculous agents against multiplying and non-multiplying microbial cells. *Am Rev Tuberc, 76*(6), 1031-1048.

Hoff, D. R., Ryan, G. J., Driver, E. R., Ssemakulu, C. C., De Groote, M. A., Basaraba, R. J., & Lenaerts, A. J. (2011). Location of intra- and extracellular M. tuberculosis populations in lungs of mice and guinea pigs during disease progression and after drug treatment. *PLoS One, 6*(3), e17550. doi:10.1371/journal.pone.0017550

Huang, Y., Lemieux, M. J., Song, J., Auer, M., & Wang, D. N. (2003). Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. *Science, 301*(5633), 616-620. doi:10.1126/science.1087619

Huard, R. C., Fabre, M., de Haas, P., Lazzarini, L. C., van Soolingen, D., Cousins, D., & Ho, J. L. (2006). Novel genetic polymorphisms that further delineate the phylogeny of the Mycobacterium tuberculosis complex. *J Bacteriol, 188*(12), 4271-4287. doi:10.1128/JB.01783-05

Hussein, M., & Mooij, J. (2002). Tuberculosis and chronic renal disease. *Saudi J Kidney Dis Transpl, 13*(3), 320-330.

Huynh, K., & Partch, C. L. (2015). Analysis of protein stability and ligand interactions by thermal shift assay. *Curr Protoc Protein Sci, 79*, 28 29 21-14. doi:10.1002/0471140864.ps2809s79

Ilari, A., Pescatori, L., Di Santo, R., Battistoni, A., Ammendola, S., Falconi, M., . . . Chiancone, E. (2016). Salmonella enterica serovar Typhimurium growth is inhibited by the concomitant binding of Zn(II) and a pyrrolyl-hydroxamate to ZnuA, the soluble component of the ZnuABC transporter. *Biochim Biophys Acta, 1860*(3), 534-541. doi:10.1016/j.bbagen.2015.12.006

Iwata, S., Ostermeier, C., Ludwig, B., & Michel, H. (1995). Structure at 2.8 A resolution of cytochrome c oxidase from Paracoccus denitrificans. *Nature, 376*(6542), 660-669. doi:10.1038/376660a0

Jackson, M., Crick, D. C., & Brennan, P. J. (2000). Phosphatidylinositol is an essential phospholipid of mycobacteria. *J Biol Chem, 275*(39), 30092-30099. doi:10.1074/jbc.M004658200

Jang, J., Becq, J., Gicquel, B., Deschavanne, P., & Neyrolles, O. (2008). Horizontally acquired genomic islands in the tubercle bacilli. *Trends Microbiol, 16*(7), 303-308. doi:10.1016/j.tim.2008.04.005

Jarlier, V., & Nikaido, H. (1994). Mycobacterial cell wall: structure and role in natural resistance to antibiotics. *FEMS Microbiol Lett, 123*(1-2), 11-18. doi:10.1111/j.1574-6968.1994.tb07194.x

Jayalakshmi, V., & Krishna, N. R. (2002). Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes. *J Magn Reson, 155*(1), 106-118. doi:10.1006/jmre.2001.2499

Jiang, D., Zhang, Q., Zheng, Q., Zhou, H., Jin, J., Zhou, W., . . . Rao, Z. (2014). Structural analysis of Mycobacterium tuberculosis ATP-binding cassette transporter subunit UgpB reveals specificity for glycerophosphocholine. *FEBS J, 281*(1), 331-341. doi:10.1111/febs.12600

Jick, S. S., Lieberman, E. S., Rahman, M. U., & Choi, H. K. (2006). Glucocorticoid use, other associated factors, and the risk of tuberculosis. *Arthritis Rheum, 55*(1), 19-26. doi:10.1002/art.21705

Jo, E. K. (2008). Mycobacterial interaction with innate receptors: TLRs, C-type lectins, and NLRs. *Curr Opin Infect Dis, 21*(3), 279-286. doi:10.1097/QCO.0b013e3282f88b5d

K. Hofmann, W. S. (1993). TMbase - A database of membrane spanning proteins segments. *Biol. Chem., 374*(166

).

Kabsch, W. (2010). XDS. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 2), 125-132. doi:10.1107/S0907444909047337

Kalscheuer, R., Weinrick, B., Veeraraghavan, U., Besra, G. S., & Jacobs, W. R., Jr. (2010). Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A, 107*(50), 21761-21766. doi:10.1073/pnas.1014642108

Kana, B. D., Gordhan, B. G., Downing, K. J., Sung, N., Vostroktunova, G., Machowski, E. E., . . . Mizrahi, V. (2008). The resuscitation-promoting factors of Mycobacterium tuberculosis are required for virulence and resuscitation from dormancy but are collectively dispensable for growth in vitro. *Mol Microbiol, 67*(3), 672-684. doi:10.1111/j.1365-2958.2007.06078.x

Kell, D. B. (2015). What would be the observable consequences if phospholipid bilayer diffusion of drugs into cells is negligible? *Trends Pharmacol Sci, 36*(1), 15-21. doi:10.1016/j.tips.2014.10.005

Kell, D. B., Dobson, P. D., Bilsland, E., & Oliver, S. G. (2013). The promiscuous binding of pharmaceutical drugs and their transporter-mediated uptake

into cells: what we (need to) know and how we can do so. *Drug Discov Today, 18*(5-6), 218-239. doi:10.1016/j.drudis.2012.11.008

Kiehn, T. E., Edwards, F. F., Brannon, P., Tsang, A. Y., Maio, M., Gold, J. W., . . . Armstrong, D. (1985). Infections caused by Mycobacterium avium complex in immunocompromised patients: diagnosis by blood culture and fecal examination, antimicrobial susceptibility tests, and morphological and seroagglutination characteristics. *J Clin Microbiol, 21*(2), 168-173.

Kiers, A., Klarenbeek, A., Mendelts, B., Van Soolingen, D., & Koeter, G. (2008). Transmission of Mycobacterium pinnipedii to humans in a zoo with marine mammals. *Int J Tuberc Lung Dis, 12*(12), 1469-1473.

Kieser, T., Moss, M. T., Dale, J. W., & Hopwood, D. A. (1986). Cloning and expression of Mycobacterium bovis BCG DNA in "Streptomyces lividans". *J Bacteriol, 168*(1), 72-80. doi:10.1128/jb.168.1.72-80.1986

Kim, M. J., Wainwright, H. C., Locketz, M., Bekker, L. G., Walther, G. B., Dittrich, C., . . . Russell, D. G. (2010). Caseation of human tuberculosis granulomas correlates with elevated host lipid metabolism. *EMBO Mol Med, 2*(7), 258-274. doi:10.1002/emmm.201000079

Koch, M., & Cote, R. (1965). Comparison of Fluorescence Microscopy with Ziehl-Neelsen Stain for Demonstration of Acid-Fast Bacilli in Smear Preparations and Tissue Sections. *Am Rev Respir Dis, 91*, 283-284. doi:10.1164/arrd.1965.91.2.283

Koch, R. (1882). The aetiology of tuberculosis. *Berliner Klinische Wochenschrift,, 19*, 221-230.

Kopp, F., Komatsu, T., Nomura, D. K., Trauger, S. A., Thomas, J. R., Siuzdak, G., . . . Cravatt, B. F. (2010). The glycerophospho metabolome and its influence on amino acid homeostasis revealed by brain metabolomics of GDE1(-/-) mice. *Chem Biol, 17*(8), 831-840. doi:10.1016/j.chembiol.2010.06.009

Korepanova, A., Gao, F. P., Hua, Y., Qin, H., Nakamoto, R. K., & Cross, T. A. (2005). Cloning and expression of multiple integral membrane proteins from Mycobacterium tuberculosis in Escherichia coli. *Protein Sci, 14*(1), 148-158. doi:10.1110/ps.041022305

Korepanova, A., Pereda-Lopez, A., Solomon, L. R., Walter, K. A., Lake, M. R., Bianchi, B. R., . . . Chiu, M. L. (2009). Expression and purification of human TRPV1 in baculovirus-infected insect cells for structural studies. *Protein Expr Purif, 65*(1), 38-50. doi:10.1016/j.pep.2008.12.006

Koul, A., Arnoult, E., Lounis, N., Guillemont, J., & Andries, K. (2011). The challenge of new drug discovery for tuberculosis. *Nature, 469*(7331), 483-490. doi:10.1038/nature09657

Kozlikova, B., Sebestova, E., Sustr, V., Brezovsky, J., Strnad, O., Daniel, L., . . . Sochor, J. (2014). CAVER Analyst 1.0: graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics, 30*(18), 2684-2685. doi:10.1093/bioinformatics/btu364

Krause. (1928). Tuberculosis and Public Health. *American Review of Tuberculosis, 18*(3), 271-322

.

Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol, 372*(3), 774-797. doi:10.1016/j.jmb.2007.05.022

Kruh, N. A., Troudt, J., Izzo, A., Prenni, J., & Dobos, K. M. (2010). Portrait of a pathogen: the Mycobacterium tuberculosis proteome in vivo. *PLoS One, 5*(11), e13938. doi:10.1371/journal.pone.0013938

Kumar, P., Kesari, P., Kokane, S., Ghosh, D. K., Kumar, P., & Sharma, A. K. (2019). Crystal structures of a putative periplasmic cystine-binding protein from Candidatus Liberibacter asiaticus: insights into an adapted mechanism of ligand binding. *FEBS J, 286*(17), 3450-3472. doi:10.1111/febs.14921

Lack, N., Lowe, E. D., Liu, J., Eltis, L. D., Noble, M. E., Sim, E., & Westwood, I. M. (2008). Structure of HsaD, a steroid-degrading hydrolase, from Mycobacterium tuberculosis. *Acta Crystallogr Sect F Struct Biol Cryst Commun, 64*(Pt 1), 2-7. doi:10.1107/S1744309107065931

Lammertyn, E., & Anne, J. (1998). Modifications of Streptomyces signal peptides and their effects on protein production and secretion. *FEMS Microbiol Lett, 160*(1), 1-10. doi:10.1111/j.1574-6968.1998.tb12882.x

Lanzetta, P. A., Alvarez, L. J., Reinach, P. S., & Candia, O. A. (1979). An improved assay for nanomole amounts of inorganic phosphate. *Anal Biochem, 100*(1), 95-97. doi:10.1016/0003-2697(79)90115-5

Larrouy-Maumus, G., Biswas, T., Hunt, D. M., Kelly, G., Tsodikov, O. V., & de Carvalho, L. P. (2013). Discovery of a glycerol 3-phosphate phosphatase reveals glycerophospholipid polar head recycling in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A, 110*(28), 11320-11325. doi:10.1073/pnas.1221597110

Larson, T. J., Ehrmann, M., & Boos, W. (1983). Periplasmic glycerophosphodiester phosphodiesterase of Escherichia coli, a new enzyme of the glp regulon. *J Biol Chem, 258*(9), 5428-5432.

Lau, G. H., MacGillivray, R. T., & Murphy, M. E. (2004). Characterization of a nucleotide-binding domain associated with neisserial iron transport. *J Bacteriol, 186*(10), 3266-3269. doi:10.1128/jb.186.10.3266-3269.2004

Laughon, B. E., & Nacy, C. A. (2017). Tuberculosis - drugs in the 2016 development pipeline. *Nat Rev Dis Primers, 3*, 17015. doi:10.1038/nrdp.2017.15

Lawson, D. M., Artymiuk, P. J., Yewdall, S. J., Smith, J. M., Livingstone, J. C., Treffry, A., . . . et al. (1991). Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature, 349*(6309), 541-544. doi:10.1038/349541a0

Le Chevalier, F., Cascioferro, A., Frigui, W., Pawlik, A., Boritsch, E. C., Bottai, D., . . . Brosch, R. (2015). Revisiting the role of phospholipases C in virulence and the lifecycle of Mycobacterium tuberculosis. *Sci Rep, 5*, 16918. doi:10.1038/srep16918

Lederer, E., Adam, A., Ciorbaru, R., Petit, J. F., & Wietzerbin, J. (1975). Cell walls of Mycobacteria and related organisms; chemistry and immunostimulant properties. *Mol Cell Biochem, 7*(2), 87-104. doi:10.1007/bf01792076

Lee, & Ballou, C. (1964). Structural Studies on the Myo-Inositol Mannodides from the Glycolipids of Mycobacterium Tuberculosis and Mycobacterium Phlei. *J Biol Chem, 239*, 1316-1327.

Lee, J. (2015). Diagnosis and treatment of extrapulmonary tuberculosis. *Tuberc Respir Dis (Seoul), 78*(2), 47-55. doi:10.4046/trd.2015.78.2.47

Lee, Y. H., Dorwart, M. R., Hazlett, K. R., Deka, R. K., Norgard, M. V., Radolf, J. D., & Hasemann, C. A. (2002). The crystal structure of Zn(II)-free Treponema pallidum TroA, a periplasmic metal-binding protein, reveals a closed conformation. *J Bacteriol, 184*(8), 2300-2304. doi:10.1128/jb.184.8.2300-2304.2002

Lehmann, J. (1946). Para-aminosalicylic acid in the treatment of tuberculosis. *Lancet, 1*(6384), 15. doi:10.1016/s0140-6736(46)91185-3

Lemieux, M. J., Huang, Y., & Wang, D. N. (2004). Glycerol-3-phosphate transporter of Escherichia coli: structure, function and regulation. *Res Microbiol, 155*(8), 623-629. doi:10.1016/j.resmic.2004.05.016

Leventis, P. A., & Grinstein, S. (2010). The distribution and function of phosphatidylserine in cellular membranes. *Annu Rev Biophys, 39*, 407-427. doi:10.1146/annurev.biophys.093008.131234

Libardo, M., Boshoff, H. I., & Barry, C. E., 3rd. (2018). The present state of the tuberculosis drug development pipeline. *Curr Opin Pharmacol, 42*, 81-94. doi:10.1016/j.coph.2018.08.001

Liberek, K., Skowyra, D., Zylicz, M., Johnson, C., & Georgopoulos, C. (1991). The Escherichia coli DnaK chaperone, the 70-kDa heat shock protein eukaryotic equivalent, changes conformation upon ATP hydrolysis, thus triggering its dissociation from a bound target protein. *J Biol Chem, 266*(22), 14491-14496.

Lin, P. L., Pawar, S., Myers, A., Pegu, A., Fuhrman, C., Reinhart, T. A., . . . Flynn, J. L. (2006). Early events in Mycobacterium tuberculosis infection in cynomolgus macaques. *Infect Immun, 74*(7), 3790-3803. doi:10.1128/IAI.00064-06

Liu, J., Bruhn, D. F., Lee, R. B., Zheng, Z., Janusic, T., Scherbakov, D., . . . Lee, R. E. (2017). Structure-Activity Relationships of Spectinamide Antituberculosis Agents: A Dissection of Ribosomal Inhibition and Native Efflux Avoidance Contributions. *ACS Infect Dis, 3*(1), 72-88. doi:10.1021/acsinfecdis.6b00158

Locher, K. P. (2009). Review. Structure and mechanism of ATP-binding cassette transporters. *Philos Trans R Soc Lond B Biol Sci, 364*(1514), 239-245. doi:10.1098/rstb.2008.0125

Locher, K. P. (2016). Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat Struct Mol Biol, 23*(6), 487-493. doi:10.1038/nsmb.3216

Loebel, S., Richardson. (1930). The respiratory metabolism of the tubercle bacillus. *Transactions National Tuberculosis Association, 26*, 196-199.

Long, E. R., & Ferebee, S. H. (1950). A controlled investigation of streptomycin treatment in pulmonary tuberculosis. *Public Health Rep, 65*(44), 1421-1451.

Loudon, R. G., Bumgarner, L. R., Lacy, J., & Coffman, G. K. (1969). Aerial transmission of mycobacteria. *Am Rev Respir Dis, 100*(2), 165-171. doi:10.1164/arrd.1969.100.2.165

Luo, H., Benner, R., Long, R. A., & Hu, J. (2009). Subcellular localization of marine bacterial alkaline phosphatases. *Proc Natl Acad Sci U S A, 106*(50), 21219-21223. doi:10.1073/pnas.0907586106

Lyon, R. H., Lichstein, H. C., & Hall, W. H. (1963). Effect of Tween 80 on the Growth of Tubercle Bacilli in Aerated Cultures. *J Bacteriol, 86*, 280-284.

Lyumkis, D. (2019). Challenges and opportunities in cryo-EM single-particle analysis. *J Biol Chem, 294*(13), 5181-5197. doi:10.1074/jbc.REV118.005602

Maartens, G., & Benson, C. A. (2015). Linezolid for Treating Tuberculosis: A Delicate Balancing Act. *EBioMedicine, 2*(11), 1568-1569. doi:10.1016/j.ebiom.2015.10.014

Mahairas, G. G., Sabo, P. J., Hickey, M. J., Singh, D. C., & Stover, C. K. (1996). Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. *J Bacteriol, 178*(5), 1274-1282. doi:10.1128/jb.178.5.1274-1282.1996

Majorek, K. A., Kuhn, M. L., Chruszcz, M., Anderson, W. F., & Minor, W. (2014). Double trouble-Buffer selection and His-tag presence may be responsible

for nonreproducibility of biomedical experiments. *Protein Sci, 23*(10), 1359-1368. doi:10.1002/pro.2520

Manabe, Y. C., Dannenberg, A. M., Jr., Tyagi, S. K., Hatem, C. L., Yoder, M., Woolwine, S. C., . . . Bishai, W. R. (2003). Different strains of Mycobacterium tuberculosis cause various spectrums of disease in the rabbit model of tuberculosis. *Infect Immun, 71*(10), 6004-6011. doi:10.1128/iai.71.10.6004-6011.2003

Manchester, K. (1984). Tuberculosis and leprosy in antiquity: an interpretation. *Med Hist, 28*(2), 162-173. doi:10.1017/s0025727300035705

Manjunatha, U., Boshoff, H. I., & Barry, C. E. (2009). The mechanism of action of PA-824: Novel insights from transcriptional profiling. *Commun Integr Biol, 2*(3), 215-218. doi:10.4161/cib.2.3.7926

Manson, A. L., Abeel, T., Galagan, J. E., Sundaramurthi, J. C., Salazar, A., Gehrmann, T., . . . Earl, A. M. (2017). Mycobacterium tuberculosis Whole Genome Sequences From Southern India Suggest Novel Resistance Mechanisms and the Need for Region-Specific Diagnostics. *Clin Infect Dis, 64*(11), 1494-1501. doi:10.1093/cid/cix169

Mao, B., Pear, M. R., McCammon, J. A., & Quiocho, F. A. (1982). Hinge-bending in L-arabinose-binding protein. The "Venus's-flytrap" model. *J Biol Chem, 257*(3), 1131-1133.

Marrakchi, H., Laneelle, G., & Quemard, A. (2000). InhA, a target of the antituberculous drug isoniazid, is involved in a mycobacterial fatty acid elongation system, FAS-II. *Microbiology, 146 (Pt 2)*, 289-296. doi:10.1099/00221287-146-2-289

Mayer, M., & Meyer, B. (1999). Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew Chem Int Ed Engl, 38*(12), 1784-1788. doi:10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *J Appl Crystallogr, 40*(Pt 4), 658-674. doi:10.1107/S0021889807021206

McCune, R. M., Feldmann, F. M., Lambert, H. P., & McDermott, W. (1966). Microbial persistence. I. The capacity of tubercle bacilli to survive sterilization in mouse tissues. *J Exp Med, 123*(3), 445-468. doi:10.1084/jem.123.3.445

McCune, R. M., Jr., McDermott, W., & Tompsett, R. (1956). The fate of Mycobacterium tuberculosis in mouse tissues as determined by the microbial enumeration technique. II. The conversion of tuberculous infection to the latent state by the administration of pyrazinamide and a companion drug. *J Exp Med, 104*(5), 763-802. doi:10.1084/jem.104.5.763

McKinney, J. D., Honer zu Bentrup, K., Munoz-Elias, E. J., Miczak, A., Chen, B., Chan, W. T., . . . Russell, D. G. (2000). Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature, 406*(6797), 735-738. doi:10.1038/35021074

Medeiros, M. A., Dellagostin, O. A., Armoa, G. R., Degrave, W. M., De Mendonca-Lima, L., Lopes, M. Q., . . . McIntosh, D. (2002). Comparative evaluation of Mycobacterium vaccae as a surrogate cloning host for use in the study of mycobacterial genetics. *Microbiology, 148*(Pt 7), 1999-2009. doi:10.1099/00221287-148-7-1999

Melief, E., Kokoczka, R., Files, M., Bailey, M. A., Alling, T., Li, H., . . . Parish, T. (2018). Construction of an overexpression library for Mycobacterium tuberculosis. *Biol Methods Protoc, 3*(1), bpy009. doi:10.1093/biomethods/bpy009

Middlebrook, G., & Cohn, M. L. (1958). Bacteriology of tuberculosis: laboratory methods. *Am J Public Health Nations Health, 48*(7), 844-853. doi:10.2105/ajph.48.7.844

Middlebrook, G., Cohn, M. L., Schaefer, W. B., & Kovitz, C. (1954). Developments in the laboratory: the growth requirements and pathogenic properties of isoniazid-resistant tubercle bacilli. *Trans Annu Meet Natl Tuberc Assoc, 96*(50th Meeting), 239-240.

Minnikin, D. E., Minnikin, S. M., Parlett, J. H., Goodfellow, M., & Magnusson, M. (1984). Mycolic acid patterns of some species of Mycobacterium. *Arch Microbiol, 139*(2-3), 225-231.

Monaco, S., Tailford, L. E., Juge, N., & Angulo, J. (2017). Differential Epitope Mapping by STD NMR Spectroscopy To Reveal the Nature of Protein-Ligand Contacts. *Angew Chem Int Ed Engl, 56*(48), 15289-15293. doi:10.1002/anie.201707682

Monot, M., Honore, N., Garnier, T., Zidane, N., Sherafi, D., Paniz-Mondolfi, A., . . . Cole, S. T. (2009). Comparative genomic and phylogeographic analysis of Mycobacterium leprae. *Nat Genet, 41*(12), 1282-1289. doi:10.1038/ng.477

Morse, D., Brothwell, D. R., & Ucko, P. J. (1964). Tuberculosis in Ancient Egypt. *Am Rev Respir Dis, 90*, 524-541. doi:10.1164/arrd.1964.90.4.524

Mostowy, S., Cousins, D., Brinkman, J., Aranaz, A., & Behr, M. A. (2002). Genomic deletions suggest a phylogeny for the Mycobacterium tuberculosis complex. *J Infect Dis, 186*(1), 74-80. doi:10.1086/341068

MRC. (1962). Medical Research Council/Tuberculosis Chemotherapy Trials Committee. Long-term chemotherapy in the treatment of chronic pulmonary tuberculosis with cavitation. *Tubercle, 43*, 201-267.

Mycobrowser. (2019). https://mycobrowser.epfl.ch.

Nakamura. (1938). Uber die Atmung der Tuberkelbazillen. *J. Exp. Med*(34), 234-245.

Nandakumar, M., Prosser, G. A., de Carvalho, L. P., & Rhee, K. (2015). Metabolomics of Mycobacterium tuberculosis. *Methods Mol Biol, 1285*, 105-115. doi:10.1007/978-1-4939-2450-9_6

Nepravishta, R., Walpole, S., Tailford, L., Juge, N., & Angulo, J. (2019). Deriving Ligand Orientation in Weak Protein-Ligand Complexes by DEEP-STD NMR Spectroscopy in the Absence of Protein Chemical-Shift Assignment. *Chembiochem, 20*(3), 340-344. doi:10.1002/cbic.201800568

Nguyen, P. T., Lai, J. Y., Lee, A. T., Kaiser, J. T., & Rees, D. C. (2018). Noncanonical role for the binding protein in substrate uptake by the MetNI methionine ATP Binding Cassette (ABC) transporter. *Proc Natl Acad Sci U S A, 115*(45), E10596-E10604. doi:10.1073/pnas.1811003115

Niederweis, M. (2008). Nutrient acquisition by mycobacteria. *Microbiology, 154*(Pt 3), 679-692. doi:10.1099/mic.0.2007/012872-0

Niesen, F. H., Berglund, H., & Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc, 2*(9), 2212-2221. doi:10.1038/nprot.2007.321

Nikaido, K., Liu, P. Q., & Ames, G. F. (1997). Purification and characterization of HisP, the ATP-binding subunit of a traffic ATPase (ABC transporter), the histidine permease of Salmonella typhimurium. Solubility, dimerization, and ATPase activity. *J Biol Chem, 272*(44), 27745-27752. doi:10.1074/jbc.272.44.27745

Nishihara, M., & Koga, Y. (1988). Quantitative conversion of diether or tetraether phospholipids to glycerophosphoesters by dealkylation with boron

trichloride: a tool for structural analysis of archaebacterial lipids. *J Lipid Res, 29*(3), 384-388.

Novikova, M., Metlitskaya, A., Datsenko, K., Kazakov, T., Kazakov, A., Wanner, B., & Severinov, K. (2007). The Escherichia coli Yej transporter is required for the uptake of translation inhibitor microcin C. *J Bacteriol, 189*(22), 8361-8365. doi:10.1128/JB.01028-07

Novy, S. (1925). Microbic respiration. II. Respiration of the tubercle bacillus. *J. Infect. Dis, 36*, 168-232.

Ohshima, N., Yamashita, S., Takahashi, N., Kuroishi, C., Shiro, Y., & Takio, K. (2008). Escherichia coli cytosolic glycerophosphodiester phosphodiesterase (UgpQ) requires Mg2+, Co2+, or Mn2+ for its enzyme activity. *J Bacteriol, 190*(4), 1219-1223. doi:10.1128/JB.01223-07

Omasits, U., Ahrens, C. H., Muller, S., & Wollscheid, B. (2014). Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics, 30*(6), 884-886. doi:10.1093/bioinformatics/btt607

Oswald, C., Smits, S. H., Hoing, M., Sohn-Bosser, L., Dupont, L., Le Rudulier, D., . . . Bremer, E. (2008). Crystal structures of the choline/acetylcholine substrate-binding protein ChoX from Sinorhizobium meliloti in the liganded and unliganded-closed states. *J Biol Chem, 283*(47), 32848-32859. doi:10.1074/jbc.M806021200

Otto, D. M., Campanero-Rhodes, M. A., Karamanska, R., Powell, A. K., Bovin, N., Turnbull, J. E., . . . Crocker, P. R. (2011). An expression system for screening of proteins for glycan and protein interactions. *Anal Biochem, 411*(2), 261-270. doi:10.1016/j.ab.2010.12.036

Palamini, M., Canciani, A., & Forneris, F. (2016). Identifying and Visualizing Macromolecular Flexibility in Structural Biology. *Front Mol Biosci, 3*, 47. doi:10.3389/fmolb.2016.00047

Pandey, S., Modak, A., Phale, P. S., & Bhaumik, P. (2016). High Resolution Structures of Periplasmic Glucose-binding Protein of Pseudomonas putida CSV86 Reveal Structural Basis of Its Substrate Specificity. *J Biol Chem, 291*(15), 7844-7857. doi:10.1074/jbc.M115.697268

Pao, S. S., Paulsen, I. T., & Saier, M. H., Jr. (1998). Major facilitator superfamily. *Microbiol Mol Biol Rev, 62*(1), 1-34.

Parent, L. J., Salam, M. M., Appelbaum, P. C., & Dossett, J. H. (1995). Disseminated Mycobacterium marinum infection and bacteremia in a child with severe combined immunodeficiency. *Clin Infect Dis, 21*(5), 1325-1327. doi:10.1093/clinids/21.5.1325

Patton-Vogt, J. L., & Henry, S. A. (1998). GIT1, a gene encoding a novel transporter for glycerophosphoinositol in Saccharomyces cerevisiae. *Genetics, 149*(4), 1707-1715.

Paula, S., Volkov, A. G., Van Hoek, A. N., Haines, T. H., & Deamer, D. W. (1996). Permeation of protons, potassium ions, and small polar molecules through phospholipid bilayers as a function of membrane thickness. *Biophys J, 70*(1), 339-348. doi:10.1016/S0006-3495(96)79575-9

Paulson, T. (2013). Epidemiology: A mortal foe. *Nature, 502*(7470), S2-3. doi:10.1038/502S2a

Peroutka Iii, R. J., Orcutt, S. J., Strickler, J. E., & Butt, T. R. (2011). SUMO fusion technology for enhanced protein expression and purification in prokaryotes and eukaryotes. *Methods Mol Biol, 705*, 15-30. doi:10.1007/978-1-61737-967-3_2

Pethe, K., Bifani, P., Jang, J., Kang, S., Park, S., Ahn, S., . . . Kim, J. (2013). Discovery of Q203, a potent clinical candidate for the treatment of tuberculosis. *Nat Med, 19*(9), 1157-1160. doi:10.1038/nm.3262

Petit, J. F., Adam, A., Wietzerbin-Falszpan, J., Lederer, E., & Ghuysen, J. M. (1969). Chemical structure of the cell wall of Mycobacterium smegmatis. I. Isolation and partial characterization of the peptidoglycan. *Biochem Biophys Res Commun, 35*(4), 478-485. doi:10.1016/0006-291x(69)90371-4

Pflugrath, J. W. (2015). Practical macromolecular cryocrystallography. *Acta Crystallogr F Struct Biol Commun, 71*(Pt 6), 622-642. doi:10.1107/S2053230X15008304

Phillips, G. N., Jr., Mahajan, V. K., Siu, A. K., & Quiocho, F. A. (1976). Structure of L-arabinose-binding protein from Escherichia coli at 5 A resolution and preliminary results at 3.5 A. *Proc Natl Acad Sci U S A, 73*(7), 2186-2190. doi:10.1073/pnas.73.7.2186

Pieters, J. (2008). Mycobacterium tuberculosis and the macrophage: maintaining a balance. *Cell Host Microbe, 3*(6), 399-407. doi:10.1016/j.chom.2008.05.006

Pitarque, S., Larrouy-Maumus, G., Payre, B., Jackson, M., Puzo, G., & Nigou, J. (2008). The immunomodulatory lipoglycans, lipoarabinomannan and lipomannan, are exposed at the mycobacterial cell surface. *Tuberculosis (Edinb), 88*(6), 560-565. doi:10.1016/j.tube.2008.04.002

Pletzer, D., Lafon, C., Braun, Y., Kohler, T., Page, M. G., Mourez, M., & Weingart, H. (2014). High-throughput screening of dipeptide utilization mediated by the ABC transporter DppBCDF and its substrate-binding proteins DppA1-A5 in Pseudomonas aeruginosa. *PLoS One, 9*(10), e111311. doi:10.1371/journal.pone.0111311

Pollock, J. M., & Neill, S. D. (2002). Mycobacterium bovis infection and tuberculosis in cattle. *Vet J, 163*(2), 115-127.

Porvaznik, I., Solovic, I., & Mokry, J. (2017). Non-Tuberculous Mycobacteria: Classification, Diagnostics, and Therapy. *Adv Exp Med Biol, 944*, 19-25. doi:10.1007/5584_2016_45

Poulsen, C., Akhter, Y., Jeon, A. H., Schmitt-Ulms, G., Meyer, H. E., Stefanski, A., . . . Song, Y. H. (2010). Proteome-wide identification of mycobacterial pupylation targets. *Mol Syst Biol, 6*, 386. doi:10.1038/msb.2010.39

Prasad, S., Khadatare, P. B., & Roy, I. (2011). Effect of chemical chaperones in improving the solubility of recombinant proteins in Escherichia coli. *Appl Environ Microbiol, 77*(13), 4603-4609. doi:10.1128/AEM.05259-11

Preiss, L., Langer, J. D., Yildiz, O., Eckhardt-Strelau, L., Guillemont, J. E., Koul, A., & Meier, T. (2015). Structure of the mycobacterial ATP synthase Fo rotor ring in complex with the anti-TB drug bedaquiline. *Sci Adv, 1*(4), e1500106. doi:10.1126/sciadv.1500106

Pulido, N. O., Silva, D. A., Tellez, L. A., Perez-Hernandez, G., Garcia-Hernandez, E., Sosa-Peinado, A., & Fernandez-Velasco, D. A. (2015). On the molecular basis of the high affinity binding of basic amino acids to LAOBP, a periplasmic binding protein from Salmonella typhimurium. *J Mol Recognit, 28*(2), 108-116. doi:10.1002/jmr.2434

Quiocho, F. A., & Ledvina, P. S. (1996). Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol Microbiol, 20*(1), 17-25.

Quiocho, F. A., Phillips, G. N., Jr., Parsons, R. G., & Hogg, R. W. (1974). Letter: Crystallographic data of an L-arabinose-binding protein from Escherichia coli. *J Mol Biol, 86*(2), 491-493. doi:10.1016/0022-2836(74)90032-1

Rabhi-Essafi, I., Sadok, A., Khalaf, N., & Fathallah, D. M. (2007). A strategy for high-level expression of soluble and functional human interferon alpha as a GST-fusion protein in E. coli. *Protein Eng Des Sel, 20*(5), 201-209. doi:10.1093/protein/gzm012

Radhakrishnan, A., Furze, C. M., Ahangar, M. S., & Fullam, E. (2018). A GFP-strategy for efficient recombinant protein overexpression and purification in Mycobacterium smegmatis. *RSC Adv, 8*(58), 33087-33095. doi:10.1039/c8ra06237d

Ramos-Gonzalez, M. I., Campos, M. J., & Ramos, J. L. (2005). Analysis of Pseudomonas putida KT2440 gene expression in the maize rhizosphere: in vivo [corrected] expression technology capture and identification of root-activated promoters. *J Bacteriol, 187*(12), 4033-4041. doi:10.1128/JB.187.12.4033-4041.2005

RCSB. (2019). https://www.rcsb.org.

Rebuffat, S. (2012). Microcins in action: amazing defence strategies of Enterobacteria. *Biochem Soc Trans, 40*(6), 1456-1462. doi:10.1042/BST20120183

Rengarajan, J., Bloom, B. R., & Rubin, E. J. (2005). Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages. *Proc Natl Acad Sci U S A, 102*(23), 8327-8332. doi:10.1073/pnas.0503272102

Robitzek, E. H., Selikoff, I. J., & Ornstein, G. G. (1952). Chemotherapy of human tuberculosis with hydrazine derivatives of isonicotinic acid; preliminary report of representative cases. *Q Bull Sea View Hosp, 13*(1), 27-51.

Rogall, T., Wolters, J., Flohr, T., & Bottger, E. C. (1990). Towards a phylogeny and definition of species at the molecular level within the genus Mycobacterium. *Int J Syst Bacteriol, 40*(4), 323-330. doi:10.1099/00207713-40-4-323

Romero-Rodriguez, A., Robledo-Casados, I., & Sanchez, S. (2015). An overview on transcriptional regulators in Streptomyces. *Biochim Biophys Acta, 1849*(8), 1017-1039. doi:10.1016/j.bbagrm.2015.06.007

Ronning, D. R., Klabunde, T., Besra, G. S., Vissa, V. D., Belisle, J. T., & Sacchettini, J. C. (2000). Crystal structure of the secreted form of antigen 85C reveals potential targets for mycobacterial drugs and vaccines. *Nat Struct Biol, 7*(2), 141-146. doi:10.1038/72413

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol, 5*, 172. doi:10.3389/fmicb.2014.00172

Rue-Albrecht, K., Magee, D. A., Killick, K. E., Nalpas, N. C., Gordon, S. V., & MacHugh, D. E. (2014). Comparative functional genomics and the bovine macrophage response to strains of the mycobacterium genus. *Front Immunol, 5*, 536. doi:10.3389/fimmu.2014.00536

Russell, D. G. (2011). Mycobacterium tuberculosis and the intimate discourse of a chronic infection. *Immunol Rev, 240*(1), 252-268. doi:10.1111/j.1600-065X.2010.00984.x

Sahu, S., & Lynn, W. S. (1977). Lipid composition of human alveolar macrophages. *Inflammation, 2*(2), 83-91.

Saida, F., Uzan, M., Odaert, B., & Bontems, F. (2006). Expression of highly toxic genes in E. coli: special strategies and genetic tools. *Curr Protein Pept Sci, 7*(1), 47-56.

Saitoh, H., Pu, R. T., & Dasso, M. (1997). SUMO-1: wrestling with a new ubiquitin-related modifier. *Trends Biochem Sci, 22*(10), 374-376.

Salo, W. L., Aufderheide, A. C., Buikstra, J., & Holcomb, T. A. (1994). Identification of Mycobacterium tuberculosis DNA in a pre-Columbian Peruvian mummy. *Proc Natl Acad Sci U S A, 91*(6), 2091-2094. doi:10.1073/pnas.91.6.2091

Sanger, I. (2019). https://www.sanger.ac.uk/resources/downloads/bacteria/mycobacterium.html.

Sassetti, C. M., Boyd, D. H., & Rubin, E. J. (2001). Comprehensive identification of conditionally essential genes in mycobacteria. *Proc Natl Acad Sci U S A, 98*(22), 12712-12717. doi:10.1073/pnas.231275498

Sassetti, C. M., & Rubin, E. J. (2003). Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A, 100*(22), 12989-12994. doi:10.1073/pnas.2134250100

Schafer, F., Seip, N., Maertens, B., Block, H., & Kubicek, J. (2015). Purification of GST-Tagged Proteins. *Methods Enzymol, 559*, 127-139. doi:10.1016/bs.mie.2014.11.005

Schatz, Bugie, & Waksman. (1944). Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. *Proc Soc Exp Biol Med, 55*, 66-69.

Scheepers, G. H., Lycklama, A. N. J. A., & Poolman, B. (2016). An updated structural classification of substrate-binding proteins. *FEBS Lett, 590*(23), 4393-4401. doi:10.1002/1873-3468.12445

Schlegel, S., Hjelm, A., Baumgarten, T., Vikstrom, D., & de Gier, J. W. (2014). Bacterial-based membrane protein production. *Biochim Biophys Acta, 1843*(8), 1739-1749. doi:10.1016/j.bbamcr.2013.10.023

Schlegel, S., Lofblom, J., Lee, C., Hjelm, A., Klepsch, M., Strous, M., . . . de Gier, J. W. (2012). Optimizing membrane protein overexpression in the Escherichia coli strain Lemo21(DE3). *J Mol Biol, 423*(4), 648-659. doi:10.1016/j.jmb.2012.07.019

Schmidl, S. R., Otto, A., Lluch-Senar, M., Pinol, J., Busse, J., Becher, D., & Stulke, J. (2011). A trigger enzyme in Mycoplasma pneumoniae: impact of the glycerophosphodiesterase GlpQ on virulence and gene expression. *PLoS Pathog, 7*(9), e1002263. doi:10.1371/journal.ppat.1002263

Schnappinger, D., Ehrt, S., Voskuil, M. I., Liu, Y., Mangan, J. A., Monahan, I. M., . . . Schoolnik, G. K. (2003). Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages: Insights into the Phagosomal Environment. *J Exp Med, 198*(5), 693-704. doi:10.1084/jem.20030846

Schneider, B. E., Behrends, J., Hagens, K., Harmel, N., Shayman, J. A., & Schaible, U. E. (2014). Lysosomal phospholipase A2: a novel player in host immunity to Mycobacterium tuberculosis. *Eur J Immunol, 44*(8), 2394-2404. doi:10.1002/eji.201344383

Schneider, E., Linde, M., & Tebbe, S. (1995). Functional purification of a bacterial ATP-binding cassette transporter protein (MalK) from the cytoplasmic fraction of an overproducing strain. *Protein Expr Purif, 6*(1), 10-14. doi:10.1006/prep.1995.1002

Schue, M., Maurin, D., Dhouib, R., Bakala N'Goma, J. C., Delorme, V., Lambeau, G., . . . Canaan, S. (2010). Two cutinase-like proteins secreted by Mycobacterium tuberculosis show very different lipolytic activities reflecting their physiological function. *FASEB J, 24*(6), 1893-1903. doi:10.1096/fj.09-144766

Seals, J. R., McDonald, J. M., Bruns, D., & Jarett, L. (1978). A sensitive and precise isotopic assay of ATPase activity. *Anal Biochem, 90*(2), 785-795. doi:10.1016/0003-2697(78)90169-0

Seung, K. J., Gelmanova, I. E., Peremitin, G. G., Golubchikova, V. T., Pavlova, V. E., Sirotkina, O. B., . . . Strelis, A. K. (2004). The effect of initial drug resistance on treatment response and acquired drug resistance during standardized short-course chemotherapy for tuberculosis. *Clin Infect Dis, 39*(9), 1321-1328. doi:10.1086/425005

Sharma, S. K., Mohan, A., Sharma, A., & Mitra, D. K. (2005). Miliary tuberculosis: new insights into an old disease. *Lancet Infect Dis, 5*(7), 415-430. doi:10.1016/S1473-3099(05)70163-8

Shigyo, K., Ocheretina, O., Merveille, Y. M., Johnson, W. D., Pape, J. W., Nathan, C. F., & Fitzgerald, D. W. (2013). Efficacy of nitazoxanide against clinical isolates of Mycobacterium tuberculosis. *Antimicrob Agents Chemother, 57*(6), 2834-2837. doi:10.1128/AAC.02542-12

Shin, S. J., Kim, S. Y., Shin, A. R., Kim, H. J., Cho, S. N., & Park, J. K. (2009). Identification of Rv2041c, a novel immunogenic antigen from Mycobacterium tuberculosis with serodiagnostic potential. *Scand J Immunol, 70*(5), 457-464. doi:10.1111/j.1365-3083.2009.02324.x

Shiver, A. L., Osadnik, H., Kritikos, G., Li, B., Krogan, N., Typas, A., & Gross, C. A. (2016). A Chemical-Genomic Screen of Neglected Antibiotics Reveals Illicit Transport of Kasugamycin and Blasticidin S. *PLoS Genet, 12*(6), e1006124. doi:10.1371/journal.pgen.1006124

Siebenlist, U., Simpson, R. B., & Gilbert, W. (1980). E. coli RNA polymerase interacts homologously with two different promoters. *Cell, 20*(2), 269-281. doi:10.1016/0092-8674(80)90613-3

Singhal, A., Arora, G., Virmani, R., Kundu, P., Khanna, T., Sajid, A., . . . Singh, Y. (2015). Systematic Analysis of Mycobacterial Acylation Reveals First Example of Acylation-mediated Regulation of Enzyme Activity of a Bacterial Phosphatase. *J Biol Chem, 290*(43), 26218-26234. doi:10.1074/jbc.M115.687269

Sita Lumsden, E. G., & Swoboda, J. A. (1952). Isoniazid in the treatment of pulmonary tuberculosis. *Tubercle, 33*(11), 322-329.

Slama, K., Chiang, C. Y., Enarson, D. A., Hassmiller, K., Fanning, A., Gupta, P., & Ray, C. (2007). Tobacco and tuberculosis: a qualitative systematic review and meta-analysis. *Int J Tuberc Lung Dis, 11*(10), 1049-1061.

Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T., & Jacobs, W. R., Jr. (1990). Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. *Mol Microbiol, 4*(11), 1911-1919.

Solovic, I., Sester, M., Gomez-Reino, J. J., Rieder, H. L., Ehlers, S., Milburn, H. J., . . . Lange, C. (2010). The risk of tuberculosis related to tumour necrosis factor antagonist therapies: a TBNET consensus statement. *Eur Respir J, 36*(5), 1185-1206. doi:10.1183/09031936.00028510

Somashekar, B. S., Amin, A. G., Rithner, C. D., Troudt, J., Basaraba, R., Izzo, A., . . . Chatterjee, D. (2011). Metabolic profiling of lung granuloma in Mycobacterium tuberculosis infected guinea pigs: ex vivo 1H magic angle spinning NMR studies. *J Proteome Res, 10*(9), 4186-4195. doi:10.1021/pr2003352

Stanley, S. A., & Cox, J. S. (2013). Host-pathogen interactions during Mycobacterium tuberculosis infections. *Curr Top Microbiol Immunol, 374*, 211-241. doi:10.1007/82_2013_332

Stinear, T. P., Seemann, T., Harrison, P. F., Jenkin, G. A., Davies, J. K., Johnson, P. D., . . . Cole, S. T. (2008). Insights from the complete genome sequence of Mycobacterium marinum on the evolution of Mycobacterium tuberculosis. *Genome Res, 18*(5), 729-741. doi:10.1101/gr.075069.107

Stover, C. K., de la Cruz, V. F., Fuerst, T. R., Burlein, J. E., Benson, L. A., Bennett, L. T., . . . et al. (1991). New use of BCG for recombinant vaccines. *Nature, 351*(6326), 456-460. doi:10.1038/351456a0

Supply, P., Marceau, M., Mangenot, S., Roche, D., Rouanet, C., Khanna, V., . . . Brosch, R. (2013). Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. *Nat Genet, 45*(2), 172-179. doi:10.1038/ng.2517

Swaim, L. E., Connolly, L. E., Volkman, H. E., Humbert, O., Born, D. E., & Ramakrishnan, L. (2006). Mycobacterium marinum infection of adult zebrafish causes caseating granulomatous tuberculosis and is moderated by adaptive immunity. *Infect Immun, 74*(11), 6108-6117. doi:10.1128/IAI.00887-06

Tabor, S., & Richardson, C. C. (1985). A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc Natl Acad Sci U S A, 82*(4), 1074-1078. doi:10.1073/pnas.82.4.1074

Tahlan, K., Wilson, R., Kastrinsky, D. B., Arora, K., Nair, V., Fischer, E., . . . Boshoff, H. I. (2012). SQ109 targets MmpL3, a membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of Mycobacterium tuberculosis. *Antimicrob Agents Chemother, 56*(4), 1797-1809. doi:10.1128/AAC.05708-11

Takayama, K., & Armstrong, E. L. (1977). Metabolic role of free mycolic acids in Mycobacterium tuberculosis. *J Bacteriol, 130*(1), 569-570.

Takayama, K., Wang, C., & Besra, G. S. (2005). Pathway to synthesis and processing of mycolic acids in Mycobacterium tuberculosis. *Clin Microbiol Rev, 18*(1), 81-101. doi:10.1128/CMR.18.1.81-101.2005

Tanaka, K. J., Song, S., Mason, K., & Pinkett, H. W. (2018). Selective substrate uptake: The role of ATP-binding cassette (ABC) importers in pathogenesis. *Biochim Biophys Acta Biomembr, 1860*(4), 868-877. doi:10.1016/j.bbamem.2017.08.011

Tang, C., Schwieters, C. D., & Clore, G. M. (2007). Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature, 449*(7165), 1078-1082. doi:10.1038/nature06232

Tang, R.-R., Yan, Z.-e., & Luo, Y.-m. (2005). Synthesis of o-L-α-glycerylphosphoryl-L-serine. *Journal of Central South University of Technology, 12*(6), 693-698.

Telenti, A., Philipp, W. J., Sreevatsan, S., Bernasconi, C., Stockbauer, K. E., Wieles, B., . . . Jacobs, W. R., Jr. (1997). The emb operon, a gene cluster of Mycobacterium tuberculosis involved in resistance to ethambutol. *Nat Med, 3*(5), 567-570.

Teng, O., Ang, C. K. E., & Guan, X. L. (2017). Macrophage-Bacteria Interactions-A Lipid-Centric Relationship. *Front Immunol, 8*, 1836. doi:10.3389/fimmu.2017.01836

Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., . . . Adams, P. D. (2008). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr, 64*(Pt 1), 61-69. doi:10.1107/S090744490705024X

ThermoFisher. (2019). https://www.thermofisher.com/order/catalog/product/89964.

Thomas, J. P., Baughn, C. O., Wilkinson, R. G., & Shepherd, R. G. (1961). A new synthetic compound with antituberculous activity in mice: ethambutol (dextro-2,2'-(ethylenediimino)-di-l-butanol). *Am Rev Respir Dis, 83*, 891-893. doi:10.1164/arrd.1961.83.6.891

Timm, J., Post, F. A., Bekker, L. G., Walther, G. B., Wainwright, H. C., Manganelli, R., . . . McKinney, J. D. (2003). Differential expression of iron-, carbon-, and oxygen-responsive mycobacterial genes in the lungs of chronically infected mice and tuberculosis patients. *Proc Natl Acad Sci U S A, 100*(24), 14321-14326. doi:10.1073/pnas.2436197100

Tischler, A. D., Leistikow, R. L., Ramakrishnan, P., Voskuil, M. I., & McKinney, J. D. (2016). Mycobacterium tuberculosis Phosphate Uptake System Component PstA2 Is Not Required for Gene Regulation or Virulence. *PLoS One, 11*(8), e0161467. doi:10.1371/journal.pone.0161467

Titgemeyer, F., Amon, J., Parche, S., Mahfoud, M., Bail, J., Schlicht, M., . . . Niederweis, M. (2007). A genomic view of sugar transport in Mycobacterium smegmatis and Mycobacterium tuberculosis. *J Bacteriol, 189*(16), 5903-5915. doi:10.1128/JB.00257-07

Tommassen, J., Eiglmeier, K., Cole, S. T., Overduin, P., Larson, T. J., & Boos, W. (1991). Characterization of two genes, glpQ and ugpQ, encoding glycerophosphoryl diester phosphodiesterases of Escherichia coli. *Mol Gen Genet, 226*(1-2), 321-327. doi:10.1007/bf00273621

Tornheim, J. A., & Dooley, K. E. (2019). The Global Landscape of Tuberculosis Therapeutics. *Annu Rev Med, 70*, 105-120. doi:10.1146/annurev-med-040717-051150

Trakhanov, S., Vyas, N. K., Luecke, H., Kristensen, D. M., Ma, J., & Quiocho, F. A. (2005). Ligand-free and -bound structures of the binding protein (LivJ) of the Escherichia coli ABC leucine/isoleucine/valine transport system: trajectory and dynamics of the interdomain rotation and ligand specificity. *Biochemistry, 44*(17), 6597-6608. doi:10.1021/bi047302o

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem, 31*(2), 455-461. doi:10.1002/jcc.21334

Udwadia, Z. F., Amale, R. A., Ajbani, K. K., & Rodrigues, C. (2012). Totally drug-resistant tuberculosis in India. *Clin Infect Dis, 54*(4), 579-581. doi:10.1093/cid/cir889

Van der Geize, R., Yam, K., Heuser, T., Wilbrink, M. H., Hara, H., Anderton, M. C., . . . Eltis, L. D. (2007). A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into Mycobacterium tuberculosis survival in macrophages. *Proc Natl Acad Sci U S A, 104*(6), 1947-1952. doi:10.1073/pnas.0605728104

van Soolingen, D., Hoogenboezem, T., de Haas, P. E., Hermans, P. W., Koedam, M. A., Teppema, K. S., . . . van Embden, J. D. (1997). A novel pathogenic taxon of the Mycobacterium tuberculosis complex, Canetti: characterization of an exceptional isolate from Africa. *Int J Syst Bacteriol, 47*(4), 1236-1245. doi:10.1099/00207713-47-4-1236

Vandal, O. H., Nathan, C. F., & Ehrt, S. (2009). Acid resistance in Mycobacterium tuberculosis. *J Bacteriol, 191*(15), 4714-4721. doi:10.1128/JB.00305-09

Veldhuizen, & Haagsman, P. (2000). Role of pulmonary surfactant components in surface film formation and dynamics. *Biochim Biophys Acta, 1467*(2), 255-270. doi:10.1016/s0005-2736(00)00256-x

Veldhuizen, E. J., & Haagsman, H. P. (2000). Role of pulmonary surfactant components in surface film formation and dynamics. *Biochim Biophys Acta, 1467*(2), 255-270. doi:10.1016/s0005-2736(00)00256-x

Veldhuizen, R., Nag, K., Orgeig, S., & Possmayer, F. (1998). The role of lipids in pulmonary surfactant. *Biochim Biophys Acta, 1408*(2-3), 90-108. doi:10.1016/s0925-4439(98)00061-1

Vergne, I., Chua, J., & Deretic, V. (2003). Tuberculosis toxin blocking phagosome maturation inhibits a novel Ca2+/calmodulin-PI3K hVPS34 cascade. *J Exp Med, 198*(4), 653-659. doi:10.1084/jem.20030527

Veyrier, F. J., Dufort, A., & Behr, M. A. (2011). The rise and fall of the Mycobacterium tuberculosis genome. *Trends Microbiol, 19*(4), 156-161. doi:10.1016/j.tim.2010.12.008

Vissa, V. D., & Brennan, P. J. (2001). The genome of Mycobacterium leprae: a minimal mycobacterial gene set. *Genome Biol, 2*(8), REVIEWS1023. doi:10.1186/gb-2001-2-8-reviews1023

Wallis, R. S., Dawson, R., Friedrich, S. O., Venter, A., Paige, D., Zhu, T., . . . Diacon, A. H. (2014). Mycobactericidal activity of sutezolid (PNU-100480) in sputum (EBA) and blood (WBA) of patients with pulmonary tuberculosis. *PLoS One, 9*(4), e94462. doi:10.1371/journal.pone.0094462

Walter, T. S., Meier, C., Assenberg, R., Au, K. F., Ren, J., Verma, A., . . . Grimes, J. M. (2006). Lysine methylation as a routine rescue strategy for protein crystallization. *Structure, 14*(11), 1617-1622. doi:10.1016/j.str.2006.09.005

Wang, F., Jain, P., Gulten, G., Liu, Z., Feng, Y., Ganesula, K., . . . Sacchettini, J. C. (2010). Mycobacterium tuberculosis dihydrofolate reductase is not a target relevant to the antitubercular activity of isoniazid. *Antimicrob Agents Chemother, 54*(9), 3776-3782. doi:10.1128/AAC.00453-10

Wartchow, C. A., Podlaski, F., Li, S., Rowan, K., Zhang, X., Mark, D., & Huang, K. S. (2011). Biosensor-based small molecule fragment screening with biolayer interferometry. *J Comput Aided Mol Des, 25*(7), 669-676. doi:10.1007/s10822-011-9439-8

Wayne, L. G., & Hayes, L. G. (1996). An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence. *Infect Immun, 64*(6), 2062-2069.

WHO. (2010). World Health Organisation: Guidelines for treatment of tuberculosis. *Fourth Edition*

.

WHO. (2014). WHO End TB Strategy.

WHO. (2018). World Health Organisation: Global Tuberculosis Report 2018

.

Wilburn, K. M., Fieweger, R. A., & VanderVen, B. C. (2018). Cholesterol and fatty acids grease the wheels of Mycobacterium tuberculosis pathogenesis. *Pathog Dis, 76*(2). doi:10.1093/femspd/fty021

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol, 112*, 531-552.

Wolber, J. M., Urbanek, B. L., Meints, L. M., Piligian, B. F., Lopez-Casillas, I. C., Zochowski, K. M., . . . Swarts, B. M. (2017). The trehalose-specific transporter LpqY-SugABC is required for antimicrobial and anti-biofilm

activity of trehalose analogues in Mycobacterium smegmatis. *Carbohydr Res, 450*, 60-66. doi:10.1016/j.carres.2017.08.003

Wuttge, S., Bommer, M., Jager, F., Martins, B. M., Jacob, S., Licht, A., . . . Schneider, E. (2012). Determinants of substrate specificity and biochemical properties of the sn-glycerol-3-phosphate ATP binding cassette transporter (UgpB-AEC2) of Escherichia coli. *Mol Microbiol, 86*(4), 908-920. doi:10.1111/mmi.12025

Xu, Y., Sugar, I. P., & Krishna, N. R. (1995). A variable target intensity-restrained global optimization (VARTIGO) procedure for determining three-dimensional structures of polypeptides from NOESY data: application to gramicidin-S. *J Biomol NMR, 5*(1), 37-48.

Yang, Y., Lee, M., & Fairn, G. D. (2018). Phospholipid subcellular localization and dynamics. *J Biol Chem, 293*(17), 6230-6240. doi:10.1074/jbc.R117.000582

Yates, M. D., & Grange, J. M. (1993). A bacteriological survey of tuberculosis due to the human tubercle bacillus (Mycobacterium tuberculosis) in south-east England: 1984-91. *Epidemiol Infect, 110*(3), 609-619. doi:10.1017/s0950268800051037

Yatsunyk, L. A., Easton, J. A., Kim, L. R., Sugarbaker, S. A., Bennett, B., Breece, R. M., . . . Rosenzweig, A. C. (2008). Structure and metal binding properties of ZnuA, a periplasmic zinc transporter from Escherichia coli. *J Biol Inorg Chem, 13*(2), 271-288. doi:10.1007/s00775-007-0320-0

Zegzouti, H., Zdanovskaia, M., Hsiao, K., & Goueli, S. A. (2009). ADP-Glo: A Bioluminescent and homogeneous ADP monitoring assay for kinases. *Assay Drug Dev Technol, 7*(6), 560-572. doi:10.1089/adt.2009.0222

Zhang, Y., Chen, C., Liu, J., Deng, H., Pan, A., Zhang, L., . . . Wan, K. (2011). Complete genome sequences of Mycobacterium tuberculosis strains CCDC5079 and CCDC5080, which belong to the Beijing family. *J Bacteriol, 193*(19), 5591-5592. doi:10.1128/JB.05452-11

Zhao, J. H., Chen, J. H., Wang, Y., Wang, Z. P., & He, Y. X. (2018). The putative compatible solute-binding protein ProX from Mycobacterium tuberculosis H37Rv: biochemical characterization and crystallographic data. *Acta Crystallogr F Struct Biol Commun, 74*(Pt 4), 231-235. doi:10.1107/S2053230X18003771

Zhao, X. (2012). Protein structure determination by solid-state NMR. *Top Curr Chem, 326*, 187-213. doi:10.1007/128_2011_287

Zimmerli, S., Edwards, S., & Ernst, J. D. (1996). Selective receptor blockade during phagocytosis does not alter the survival and growth of Mycobacterium tuberculosis in human macrophages. *Am J Respir Cell Mol Biol, 15*(6), 760-770. doi:10.1165/ajrcmb.15.6.8969271

Zimmerman, M. R. (1979). Pulmonary and osseous tuberculosis in an Egyptian mummy. *Bull N Y Acad Med, 55*(6), 604-608.